



371.25 C48s (2)  
Chapman & Rush  
Scientific measure-  
ment of classroom  
products

Acc. No. 235207  
JAN 29 1917

371.25 C48s (2)  
**KEEP YOUR CARD IN THIS POCKET**

Books will be issued only on presen-  
tation of proper library cards.

Unless labeled otherwise, books may  
be retained for four weeks. Borrowers  
finding book marked, defaced or mutilated  
are expected to report same at library  
desk; otherwise the last borrower will be  
held responsible for all imperfections  
discovered.

The card holder is responsible for all  
books drawn on his card.

No books issued unless penalties are  
paid.

Lost cards and change of residence  
must be reported promptly.



**PUBLIC LIBRARY,  
KANSAS CITY, MO.**

**KEEP YOUR CARD IN THIS POCKET**

KANSAS CITY MO PUBLIC LIBRARY







THE  
SCIENTIFIC MEASUREMENT  
OF CLASSROOM PRODUCTS

J. CROSBY CHAPMAN

B.A. (CANTAB), D.Sc. (LONDON), PH.D. (COLUMBIA)

ASSOCIATE PROFESSOR OF EXPERIMENTAL EDUCATION  
WESTERN RESERVE UNIVERSITY

GRACE PREYER RUSH

M.A. (WESTERN RESERVE UNIVERSITY)

INSTRUCTOR IN PHILOSOPHY, WESTERN RESERVE  
UNIVERSITY



SILVER, BURDETT & COMPANY

BOSTON

NEW YORK

CHICAGO

COPYRIGHT, 1917

By SILVER, BURDETT & COMPANY

## PREFACE

THE aim of this book is stated in detail in the introduction. In short, it is an attempt to present, free from its usual accompaniment of statistical methods, the new idea of educational measurement by means of objective scales. This book will not satisfy the statistically trained reader, for to such its methods will appear clumsy and inelegant. However, not every member of the teaching profession can be expected to have statistical training; yet it is essential that every one in school work have the quantitative point of view. If this small book succeeds in introducing the reader to the new movement of which the objective scales are the product, it will in a large measure make up for its obvious shortcomings.

It is our pleasure to record our obligations to the following for their courtesy in permitting us to quote their original papers and reproduce their scales:

Dr. L. P. Ayres (Handwriting and Spelling Scales).

Dr. F. W. Ballou (English Composition Scale).

Dr. B. R. Buckingham (Spelling Scales).

Mr. S. A. Courtis (Arithmetic, Writing, and Reading Scales).

Mr. W. S. Gray (Reading Scale).

Dr. M. B. Hillegas (English Composition Scale).

Dr. Daniel Starch (Reading and Spelling Scales).

Dr. E. L. Thorndike (Handwriting, Reading, and Drawing Scales).

Dr. M. R. Trabue (Language Scale).

Dr. C. Woody (Arithmetic Scale).

The full references to these original papers are given in the Appendix. We are also under obligation to Dr. H. Austin Aikins and Miss Myra E. Hills for reading parts of the manuscript.

One of the authors wishes to express his great indebtedness to Dr. E. L. Thorndike, from whose writings the ideas embodied in this book largely originated.

J. C. C.

G. P. R.

WESTERN RESERVE UNIVERSITY,  
1917.

# TABLE OF CONTENTS

	PAGE
INTRODUCTION. THE AIM OF THE BOOK . . . . .	vii
CHAPTER	
I. OBJECTIVE VERSUS SUBJECTIVE SCALES OF MEASUREMENT IN THE CLASSROOM . . . . .	1
II. SCALES FOR THE MEASUREMENT OF ABILITY IN ARITHMETIC . . . . .	9
III. SCALES FOR THE MEASUREMENT OF ABILITY IN HAND-WRITING . . . . .	31
IV. SCALES FOR THE MEASUREMENT OF ABILITY IN READING . . . . .	63
V. SCALES FOR THE MEASUREMENT OF ABILITY IN SPELLING . . . . .	107
VI. SCALES FOR THE MEASUREMENT OF ABILITY IN ENGLISH COMPOSITION . . . . .	131
VII. SCALE FOR THE MEASUREMENT OF LANGUAGE ABILITY	157
VIII. SCALE FOR THE MEASUREMENT OF ABILITY IN DRAWING . . . . .	164
IX. THE APPLICATION OF THE SCALES IN THE SCHOOLS .	173
X. DANGERS INCIDENTAL TO THE USE OF THESE SCALES	182
APPENDIX A. SOURCES FROM WHICH THE FULL ACCOUNT OF THE SCALES CAN BE OBTAINED .	189
APPENDIX B. BIBLIOGRAPHY (limited) . . . . .	191

“Are you content now?” said the Caterpillar.

“Well, I should like to be a little larger, if you wouldn’t mind,” said Alice. “Three inches is such a *wretched* height to be.”

“It is a very *good* height indeed !” said the Caterpillar angrily, rearing upright as it spoke (it was exactly three inches high).

— LEWIS CARROLL. *Alice in Wonderland.*

## INTRODUCTION

### THE AIM OF THE BOOK

IT may safely be said that the greatest contribution which has been made to education in the last ten years is the application of scientific measurement to school products. The new educational method which has resulted shows a clear recognition of the scientific spirit. That a new method was needed was generally agreed, and it has been accepted by those who have experimented with it. However, there is always danger when those engaged in the practice of a subject are for any reason unacquainted with the latest advances. This is clearly demonstrated in a field such as medicine, where it often takes years for a principle which has been accepted by the leaders in the profession to be put into practical application by the general practitioner. Particularly is this true in the field of education, for any desired advance depends on the closest coöperation of the theorist and the teaching force, largely because the classroom is the laboratory of the experimental educationalist. It is the general belief that the teacher is interested and eager to know what degree of success his efforts have won, as measured by the quality of the work done by his pupils. The difficulty which, up to the present time, has confronted the teacher has been the fact that methods for the measurement of school products have been so complicated with

statistical data that it has been almost impossible for the ordinary reader to comprehend the movement, and still more impossible for the teacher to apply these methods in the classroom. It is the object of this book to present, in a manner free from statistical data and other complicated material, a few of the more important scales which have been worked out and which can profitably be used in the course of ordinary school work by any teacher without special training. No attempt has been made to give a complete presentation of a subject which has been so recently developed, and which is still passing through the experimental stage. The authors have chosen to describe the method of construction of a few of the more important scales rather than to cover all scales which have been published up to the present time. Provided the reader becomes acquainted with the general idea of this scientific movement by using the scales here described, there is no danger that he will fail to use new scales as they are developed.

Everything has been sacrificed to clearness; even uniformity in the method of presentation has had to give way, where such uniformity would have resulted in lack of clearness.

Each scale has been presented as a unit. This has made necessary a certain amount of repetition, but the advantages of the method are apparent.



# THE SCIENTIFIC MEASUREMENT OF CLASSROOM PRODUCTS

## CHAPTER I

### OBJECTIVE VERSUS SUBJECTIVE SCALES OF MEASUREMENT

MORE and more, as we come to analyze the educational process, the old idea that this process goes on as a whole is being abandoned. In spite of obvious dangers, the more enlightened view regards the education of any particular individual as the conscious attempt of society to make that individual advance along certain desirable paths — the desirability of any particular path being determined by the capacity of the individual and the demands of society.

Along some lines the school insists that every pupil shall advance; for example, he must improve along the lines of activity which for convenience are termed arithmetic, writing, reading, English composition, spelling, drawing, etc. If we regard the pupil as advancing simultaneously along all these partly independent lines, we need not, and for many purposes should not, regard education as a single process but rather as a series of processes, each of which, when recognized, admits of study by any person who is prepared to take the time to specialize in that direction. Even such aims as the school has in regard to the building of character must, in the same way, be regarded as an attempt on the part of the school to make the pupils improve on a moral scale, which at present exists merely in thought.

If those interested in education would consistently take this analytical position, there would be a great change of attitude towards educational problems. For when we look at education as some large general process, the task of improving that process appears rather formidable; but when it is seen that general improvement is merely advance in certain specific and narrow directions desired by society, the problem of advance becomes, comparatively speaking, a simple one. There still remains the great question of what activity is the most desirable; this question must be left as one of the fundamental problems of the philosophy of education. When, however, it has been decided that a certain line of activity must be pursued in the schools, then the question and problem for the ordinary teacher is this: How can I most efficiently train the pupils to improve in skill along this line? What method should I adopt to bring the child to a reasonable efficiency in the minimum time? That is, from the point of view of the teacher there is often no doubt of what has to be done, the problem being, how can the result be accomplished with the greatest economy of effort in the minimum time. In other words, it is not so much a matter of *what* to teach, as of *what method* shall be used in teaching.

But when it comes to a choice between the various methods of teaching a given subject, what is to be the criterion? Is it to be the opinion of the teacher, or of the supervisor, or of the superintendent? If so, on what is this opinion based? In education, the time has passed when the inventor of a certain system can make unchallenged claims with regard to the success of his method; for in many subjects it is now possible to measure, under scientific conditions and independently of any individual judgment, the results obtained under various methods. In this way there is arising a science of method founded on the secure basis of accurate measurement of results.

The first essential, then, for the teacher, if sound judgment of the success of classroom instruction is to be secured, is to have some scientific methods of measuring, at intervals, the increase of skill or the improvement of the class along the lines of activity for which that teacher is responsible. Just as in a gymnasium it is possible to measure the increase in height from month to month or from year to year by means of a scale of height, so in the school the teacher must be able to measure the rate of improvement of pupils in the various subjects taught.

This really is no new idea in education; the school has always been more or less interested in measurement, as the common practice of examinations proves. It is not a question of *whether* we are to measure the efficiency of the pupils or not, but rather *how* we are to measure this efficiency. Shall the judgment be the opinion — frequently offhand or prejudiced — of some one individual; or shall the judgment be determined by the use of a standard devised by and based upon the consensus of opinion of many experts? If the results of classroom work are to be measured with any degree of exactness, then what we need are scales for measurement, scales which are as independent as possible of the judgment of the individual who uses them.

The question, therefore, of urgent importance, is: What are the present methods of measuring efficiency in the schools, and how satisfactory are these methods? It is true that there is perhaps no part of the teacher's work which he knows to be more unsatisfactory than the usual method of awarding grades and marks. Two methods are now in vogue; namely, the percentage method, and the letter mark method. In the former the pupil's work is graded on a basis of 100 as the standard; in the latter, a letter such as *E*, *G*, *F*, is given to indicate a certain degree of efficiency.

In a recent article in the *Educational Review* a well-

known writer makes the statement: "85% as a class average in subjects like arithmetic or grammar is not excessive." This statement may be true or false, but in any case it is valueless, for the simple reason that a mark of 85% never means the same standard to one individual that it does to another. In a reply to this article one writer states: "What 85% means is absolutely unknown and unknowable — *quot homines tot sententiae!*" The same argument applies to a grading indicated by letter. What guarantee is there that the same grading represents the same standard of work, when measured by different individuals? All attempts which have been made to investigate this subject prove conclusively that, even in the same school, two teachers will often give the same grading for work which is by no means equivalent. What, then, is wrong with such scales of marking? Obviously, the errors that arise from their use are due to the fact that they depend too much on the individual judgment of the teacher, or, in other words, that the scales are too *subjective*.

Opinions of teachers on handwriting form an excellent illustration of the dangers and disadvantages of subjective judgments. When a teacher says of a particular sample of writing that it is "good," "fair," "poor," not only does this judgment fail to give an absolute measure of efficiency, but even the judgment itself is largely determined by the extent to which the teacher is partial to such characteristics as legibility, grace, character, or to various styles of writing, such as slanting or vertical. In the writing scale to be described, a successful attempt has been made to eliminate this type of unscientific judgment.

In opposition to these *subjective scales* of measurement, which depend so much upon the judgment of the individual, there are scales such as those used in measuring mass, length, or time. In the use of these *objective scales*, very little depends on the judgment of the individual.

When one says that a particular body weighs 14.6 pounds or that the length of a certain rod is 18.1 feet, there is no room for dispute, since such measurements are outside the range of personal opinion. In other words, they are what we call *universal* or *objective*, for they mean the same thing to all persons at all times and in all places. On the contrary, judgments of plays, books, moral characteristics, depend very much on the character and taste of the individual. The designation "good" used by different individuals may mean very different degrees of merit; that is, the judgments are *subjective*. In the light of what we have said we may define a perfectly objective scale as a scale in respect to whose meaning all competent thinkers agree; while a perfectly subjective scale is one in respect to whose meaning all those competent to judge would be likely to disagree, save by chance.

When subjective scales, such as those described, are used in schools, it is evident that we can have no scientific basis for comparison. Yet all agree that improvement and advance depend largely upon critical comparison. Up to the present time, therefore, one of the great methods of obtaining efficiency in the outside world, has not been employed in education, because critical comparison could not safely be based on subjective judgments. When objective scales are employed in the schools, then it will be feasible to compare the work done in one school with the work done in another, or the work done under one method of instruction with the work done under a different method. Even now, in certain subjects the school administrator is able to compare teacher with teacher, school with school, system with system, and even country with country.

The great problem of measurement in education, therefore, is to construct objective or universal scales, about the use of which there can be no misunderstanding when they are placed in the hands of competent teachers.

Every such scale must fulfill at least three essential requirements: (1) It must measure a desired product; (2) it must be so simple in its application that it is suitable for ordinary classroom use; (3) it must not require an undue amount of time in administration.

From the very nature of measurement it is apparent that ability in such a subject as arithmetic admits of being measured objectively. Any competent teacher would be capable of constructing a scale to measure improvement in addition. But the essential thing is that every one shall agree to use the same method, or standard, just as they agree to use a gram, a centimeter, and a second in measuring mass, length, and time. Thus, suppose it is desired to measure speed in adding, all that is necessary is to construct a blank on which are printed the columns of figures. The test can then be administered by allowing, let us say, two minutes for the work, that is, less time than it takes even the fastest pupil to complete all the addition. Provided the same directions are followed in each case, it is possible to measure by the same standard any other school in any other system. In this way a comparison of the two groups will be perfectly easy. The essential point, then, is that all shall agree to use the same scale, under the same conditions, giving the same time allowance, and correcting and scoring in the same way. It is to fulfill these conditions that objective scales are necessary.

While arithmetic lends itself to such objective measurements, in other important subjects it is more difficult to construct scales for the measurement of efficiency, which will be relatively independent of the judgment of the teacher. It would be ideal if scales could be constructed which would measure improvement in writing, reading, drawing, English composition, spelling, etc., about the use of which there would be as little division of opinion as there is about the employment of a yardstick, a balance,

or a watch, to measure length, mass, and time, respectively. In the following pages a few of the more essential objective scales which have been worked out with this idea in view, will be presented. No claim is made that they eliminate completely the factor of the judgment of the individual teacher. Through the description and use of the scales themselves the reader may judge of the extent to which individual opinion, bias, and prejudice, as factors, have been excluded.

It will be seen that a scale may be used by the teacher merely to measure the improvement of a particular class or individual. It is advantageous, however, after a particular test has been administered, to know how the grade taking the test compares with similar grades in other school systems. In certain cases it is possible to make this comparison, for the scales have been tested with a sufficient number of pupils to establish averages of achievement, or, in other words, *norms* or *standards* for the various grades. The process of standardizing a test is quite simple. All that is necessary is to administer the test, under the identical standard conditions, in like grades, in different representative school systems, and from these results to determine the average work done in the various grades. It will be noted, from what follows, that we do not have a different scale for each grade, for in many cases all grades can be measured by the same scale. Just as we measure the dwarf and the giant with the same foot rule, and express the result in the same unit, inches, so we may measure the ability of individuals at different points of their training on the same scale, expecting of course increasing products at successive stages. In so far as standards have been established for the grades, they are included in the description given in the following chapters (II-VIII). In all cases these standards or norms must be looked upon as provisional, for none of the tests have been tried upon a sufficiently wide range of schools and

school systems to make it certain that the standards are the average achievements of the particular grade in question.

### EXERCISES

1. What are the present methods by which you measure the efficiency of your class work? Why is it so difficult to tell how your class compares with other classes of the same grade? How would such information increase your efficiency?

2. Taking twenty compositions of varying merit, grade them according to your usual method. Put them away for a month and then grade again. How do the results compare? Repeat the experiment with a series of handwriting samples.

3. Does the same mark given by different teachers imply the same standard of work on the part of the pupils? How would you prove the correctness of your answer?

4. What is the final test of any particular method of teaching a subject? Why is there so much difference of opinion with reference to the relative values of different methods?

5. What would happen if a foot meant a different length in different parts of the country? What is the effect of 80% in one school not meaning the same as 80% in another school?

6. Give ten examples of qualities which we measure (a) objectively; (b) subjectively. Is there any such thing as a perfect objective measure? Give your reasons.

7. What is meant by a norm? How would you establish the norms of height and weight of the grades in a school? How could we use the same idea in measuring growth in spelling, arithmetic, writing, and reading skill? What are the difficulties?

8. What are the disadvantages of the present system of subjective marking as they affect (a) the pupil; (b) the teacher; (c) the administration of the school system?

9. It has been shown that the same arithmetic paper received 85% and 40% when marked by two trained judges; how could this happen? What might have been done to avoid it?

10. If you had 5000 arithmetic papers, and five judges had each to mark 1000 of these, how would you attempt to secure uniformity in the system of marking?



## CHAPTER II

### ARITHMETIC SCALES

#### I. COURTIS TESTS IN ARITHMETIC

#### II. WOODY TESTS IN ARITHMETIC

WHEN it is considered how many different operations are covered by the inclusive term "arithmetic," it becomes apparent at once how little specific meaning is conveyed by the assertion that a pupil is good or poor in that subject. Since arithmetical skill is not a single ability, but consists rather of a number of abilities, discussion of it should be expressed in terms of these. For instance, instead of saying that a child is good in arithmetic, it is more accurate and far more useful to state in what specific process or processes — adding, subtracting, reasoning, etc. — he excels, for a child may be good in one of these operations and poor in another. Thus, the teacher is confronted with a problem of analysis. It must be discovered first of all in what particular process—adding, multiplying, subtracting, etc.—the pupil is weak, and then an attempt must be made to strengthen him in that process. To facilitate this work is the chief object of the Courtis Tests in Arithmetic.

Eight or nine years ago, in an early effort to measure efficiency in certain phases of arithmetical work, Courtis discovered that the ability of a given individual in some one process was very different from his ability in another. One child might be very good in addition and poor in multiplication, while another might be good in both addition and multiplication but poor in reasoning, etc. Courtis immediately began experimental work to control this individual variation; that is, to make the

child do equally well in all these operations. For several years the attempt failed; for with his increased effort at control, Courtis found that the difference in the ability of the individual in the various branches also increased. However, this work was not without most important results. As an outcome of administering the tests to more than 48,000 children in about 70 schools in 10 states, Courtis discovered one fundamental fact; namely, that one of the great factors in education is the variability of the natural abilities of children. In the first place, no child will do equally well in all the operations involved in arithmetic; for example, he may do very well in division and still do poorly in addition, or vice versa. That is, there is a difference among his special attainments or abilities in these sub-branches. Secondly, there exists a great difference in the general ability of different children. According to Courtis, these two facts mean that new educational methods, methods that will give each child a chance to develop in his own way and along his own lines, will have to be invented. In the work of analysis thus necessitated, the Courtis Tests will be of great assistance.

I. COURTIS TESTS

These original tests, called by Courtis "Series A," are eight in number and are designed to test those abilities which constitute most of that complex product known as arithmetical efficiency.

Courtis Tests — Series A

NUMBER OF TEST	FUNCTION	TIME FOR ADMINISTRA- TION OF TESTS
1	Addition	One minute
2	Subtraction	" "
3	Multiplication	" "
4	Division	" "
5	Copying Figures (Rate of motor activity)	" "
6	Speed Reasoning (Judgments of operation to be used in simple one-step problems)	" "
7	Fundamentals (Abstract examples in the four opera- tions)	Twelve minutes
8	Reasoning (Two-step problems)	Six "

These eight tests are printed on separate sheets of paper, and folders, containing full directions intended to secure uniformity of administration and marking, may be had by the examiner. For example, in the Addition Test given on page 12, the child is supposed to add across the paper from left to right. That is, his answers should be, 9, 18, 13, 8, etc. He should do as many of these problems as possible in the time allowed—one minute—and his score will be the number of problems he has done correctly in that time.

---

 Arithmetic—Test No. 1      Speed Test—Addition

Write on this paper, in the space between the lines, the answers to as many of these addition examples as possible in the time allowed.

8 9 7 8 2	1 3 6 0 3	1 7 9 3 2	1 6 9 0 4
1 9 6 0 5	5 8 9 7 2	3 7 6 0 4	2 6 5 1 2
_____	_____	_____	_____

5 8 6 9 4	1 2 5 6 7	3 4 7 0 3	1 4 8 0 2
1 3 5 0 3	4 9 8 0 2	1 6 9 8 5	6 7 9 5 7
_____	_____	_____	_____

1 8 6 0 5	4 8 9 5 3	1 3 8 2 3	2 9 7 4 5
9 4 7 2 4	1 8 7 0 6	7 9 5 0 7	2 3 8 0 2
_____	_____	_____	_____

3 7 9 0 4	2 4 5 1 6	9 2 5 0 6	7 4 8 0 3
3 4 8 6 5	1 8 9 0 2	1 8 7 4 3	1 9 6 0 4
_____	_____	_____	_____

6 9 8 1 2	1 6 7 0 2	5 9 6 7 5	4 8 5 0 7
1 4 7 1 3	8 4 5 3 6	5 2 8 0 3	4 2 6 9 3
_____	_____	_____	_____

1 4 9 0 4	1 2 6 0 3	8 9 7 8 5	1 7 9 3 2
6 7 5 1 2	6 7 9 7 2	1 9 6 0 2	3 7 6 0 4
_____	_____	_____	_____

---

The same mode of procedure applies to the Subtraction, Multiplication, and Division Tests.

In the test of reasoning with one-step problems (Test No. 6), the pupil is not required to work out the problems but merely to record what operation — addition, subtraction, etc. — he would use if he were going to work them out; this is to distinguish between skill in reasoning and mere skill in rapid calculation. In the reasoning test involving two steps (Test No. 8), the answer is to be found and recorded.

The tests just described were designed to measure the relation existing between the simpler abilities tested in the first six tests and the more complex abilities tested in the last two tests. That is, their object was to investigate whether a child who is good or poor in addition, subtraction, multiplication, or division is also good or poor, as the case may be, in fundamentals and reasoning. Courtis claims the tests have accomplished this purpose.

### Arithmetic — Test No. 7 — Fundamentals

In the blank space below, work as many of these examples as possible in the time allowed. Work them in order as numbered, writing each answer in the "answer column" before commencing a new example. Do not work on any other paper.

NO.	OPERATION	EXAMPLE	ANSWER
1	Addition	a. $25 + 830 + 122 =$ b. $232 + 8021 + 703 + 3030 =$	
2	Subtraction	a. $5496 - 163 =$ b. $943276 - 812102 =$	
3	Multiplication	$2012 \times 213 =$	
4	Division	$158664 \div 132 =$	
5	Addition	$6134 + 213 + 4800 + 6005 + 474 =$	
6	Subtraction	$73210142 - 49676378 =$	
7 } 8 }	Multiplication	$46505 \times 456 =$	
9	Division	$27217182 \div 6 =$	
10 } 11 }	Division	$3127102 \div 463 =$	
12 } 13 }	Addition	$85586 + 69685 + 39397 +$ $95836 + 37768 + 69666 +$ $78888 + 54987 =$	
14	Subtraction	$15655431 - 5878675 =$	
15 } 16 }	Multiplication	$78965 \times 678 =$	
17	Division	$44502486 \div 7 =$	
18 } 19 }	Division	$5373003 \div 769 =$	

Arithmetic — Test No. 8 — Reasoning

In the blank space below, work as many of the following examples as possible in the time allowed. Work them in order as numbered, entering each answer in the "answer column" before commencing a new example. Do not work on any other paper.

1. A party of children went from a school to a woods to gather nuts. The number found was but 205, so they bought 1,955 nuts more from a farmer. The nuts were shared equally by the children and each received 45. How many children were there in the party?

2. One summer a farmer hired 43 boys to work in an apple orchard. There were 35 trees loaded with fruit and in 57 minutes each boy had picked 49 apples. If in the beginning the total number of apples on the trees was 19,677, how many were there still to be picked?

3. A girl found by careful counting that there were 87 letters more on a page in her history than on a page of her reader. She read 31 pages in each book in the first 29 days of school. How many more letters each day did she read in one book than in the other?

4. The children of a school made small boxes to be filled with candy and given as presents at a school party. Six hundred were needed. In 4 days grades III to VII made 20, 25, 83, 150 and 150 boxes. The eighth grade agreed to make the rest. How many did the eighth grade make?

5. A girl's record in spelling for 5 days was 19, 18, 20, 16 and 20 words spelled correctly out of 20. If each of the 16 children in the grade had had the same record, what would have been the total number of words spelled correctly by that grade in 5 days?

6. A party of boys went on a long bicycle trip. They traveled 1702 miles in 37 days. A number of men then joined the party, and soon the party was traveling 58 miles per day. How much change in the number of miles ridden a day did the presence of the men make?

7. A teacher corrected 2400 arithmetic test papers; 2295 of these he marked "poor," "good" etc. All the others were marked "unsatisfactory." If each of the papers in this group had 47 mistakes, what was the total number of mistakes in the unsatisfactory papers?

8. In two schools five teachers recorded the number of blocks the children walked in going to and from the school. The total for one school was 3000 blocks; for the other 2400. The number of children in both schools was 216. How many blocks did each child walk a day?

## Courtis Tests—Series B

Courtis has recently constructed a new series of more difficult tests, "Series B," to be used in the primary grades for testing more complex operations in the four fundamental processes. The figures in these tests are chosen so that all the fundamental combinations are included.

NUMBER OF TEST	FUNCTION	TIME FOR ADMINISTRA- TION OF TESTS
1	Addition	Eight minutes
2	Subtraction	Four "
3	Multiplication	Six "
4	Division	Eight "

In the Addition Test the pupil is required to add as many figures as possible in eight minutes. In this way it may be determined whether or not a child or class has learned (1) the fundamental combinations; (2) the mechanism of column addition; (3) to carry; (4) to hold the attention; (5) to control the effects of fatigue or boredom; (6) to work at a high speed; (7) to work with accuracy. In a similar manner, each of the other three tests is put in the simplest form necessary to serve as a general measure of ability in that operation.

## Test No. 1—Addition

You will be given eight minutes to find the answers to as many of these addition examples as possible.

927	297	136	486	384	176	277	837
379	925	340	765	477	783	445	882
756	473	988	524	881	697	682	959
837	983	386	140	266	200	594	603
924	315	353	812	679	366	481	118
110	661	904	466	241	851	778	781
854	794	547	355	796	535	849	756
965	177	192	834	850	323	157	222
<u>344</u>	<u>124</u>	<u>439</u>	<u>567</u>	<u>733</u>	<u>229</u>	<u>953</u>	<u>525</u>

(Copyright by S. A. Courtis)



Series B may be used from the fourth grade up. When these four tests are standardized, which will take place as soon as more returns from their use are available, it will be possible to tell the degree of skill in each test which the average child in any particular grade should attain.

It is to be remembered that the Courtis Tests are "neither lesson sheets nor examination papers." They are only methods of investigation — mere measuring rods. By their use are revealed the actual arithmetical conditions existing in schools, classes, and individuals. To find the causes of any unsatisfactory conditions which the tests may reveal, and to remove these causes, is another problem.

The repeated use of these scales will tend to reveal the laws of development as they operate in the classroom, and to measure the efficiency of any particular educational method. A class which is being taught division by a certain method may be tested at intervals to see what improvement is taking place. If little or no improvement is shown, it may be safely inferred that the method used is not suited to that particular group; and new methods of instruction may be devised and tested, until the improvement is so marked as to leave little doubt that the method finally adopted is the one which produces the best results with the group in question. In short, the tests of Series B are scientific measures of efficiency in four operations of arithmetic, which may be used to determine the best methods of teaching these operations. Since the same tests, or their equivalents, are used in all the grades, a child or group of children may be measured over and over again, and the progress determined by the changes in the score, just as height is measured over and over again with the same measuring rod.

Since the use of standard tests makes objective scoring possible, any teacher can easily establish objective stand-

ards of work for a class; and in time it will be known what the actual standards for different school systems are. To facilitate this work of standardization, Courtis has published printed folders of instruction covering every phase of the testing, such as scoring, tabulating results, the making of graphs, etc. Very likely Series B will eventually displace Series A, except for the solution of special problems, and standards of permanent value will be obtained from its use. To those who wish to give a single test, merely to see the nature of the experiment or to measure the general character of the arithmetical work of a class as compared with that done in another class or school, the test on fundamentals (Test No. 7) in Series A is recommended, for it is a general measure of the ability to add, subtract, multiply and divide with whole numbers.

The administration of these tests is an easy matter. The twelve tests — the eight of Series A and the four of Series B — are printed on separate sheets of paper, each containing complete directions for its use. It is advisable to procure with these test sheets the manual containing full directions for the giving of the tests; for the essence of this movement lies in uniformity of administration and marking. The Courtis Standard Tests for Arithmetic may be obtained at the Department of Coöperative Research, 82 Eliot Street, Detroit, Michigan.

If a teacher desires merely to compare the general character of the work of a class with the work of other classes of the same grade, all that is necessary is to send for Test No. 7, Series A, together with the folder relating to the tests of that series. If, after the administration of this single test, more specific information is desired regarding the work of pupils in the various sub-branches — addition, subtraction, etc. — other tests in Series A, Test No. 1, for addition, Test No. 2, for subtraction, and so on, may be procured and administered. In the fourth grade and

above it is advisable to use the tests of Series B, as they are cheaper and require less time to administer.

The actual application of the test is very simple and requires but little time—from one to twelve minutes according to the test. For example, in giving Test No. 1, Series A (Addition), the teacher, after reading the instructions for administering the test as given in the manual for Series A, will proceed somewhat as follows. Holding up one of the test sheets before the pupils, the teacher will give directions for filling out the blank spaces at the top of the paper with the name of pupil, the grade, and name of school. Then, in a manner calculated to secure coöperation, the pupils will be told just what is expected of them; namely, at a given signal, "Start," to add across the paper from left to right, putting down the answers in the spaces allowed between the lines until the signal, "Stop," is given. This signal is given after one minute's time has elapsed. The teacher later records the number of problems each child has done correctly. This constitutes his score. A more or less similar course is followed with each of the other tests.

The only warning to be observed in the administration of the tests is that care must be taken to see that all the pupils start and stop at the same time, and that every effort be made to secure the interest and coöperation of the children. The work itself should proceed smoothly and steadily with no hurry or excitement. Class averages may be obtained from the record of the individual scores and such averages may be compared with those obtained from different parts of the country. (See Standard Scores.)

Within the classroom the teacher is in a position to determine which children should be selected for special attention. For example, if a child's record shows him to be very high in multiplication and low in addition, efforts should be made to improve the latter, and he

should not be made to waste time on multiplication drills. Tests administered at the beginning of school in September will show what children fall below the standard for each process in that grade. Several tests during the year will show the efficiency or inefficiency of the methods used to bring these children's records up to standard. Furthermore, a child's improvement may be followed from grade to grade by keeping a record of each pupil's score. The results obtained from the administration of such tests also make possible the accurate comparison of school systems and classes. These tests mean better work on the part of teachers because they reveal just what they are accomplishing; they mean progressive educational changes brought about through those methods of instruction which have produced the best results.

### STANDARD SCORES

As a result of administering the eight tests in Series A to almost 6700 pupils throughout the United States, Courtis has worked out the following tentative standard scores. These, it should be noted, are the average scores actually obtained by the pupils themselves.

Test No.	1	2	3 & 4	5	No. 6		No. 7		No. 8	
					Atts.	Rts.	Atts.	Rts.	Atts.	Rts.
Grade III . . .	26	19	16	58	2.7	2.1	5.0	2.7	2.0	1.1
Grade IV . . .	34	25	23	72	3.7	3.0	7.0	3.3	2.6	1.7
Grade V . . .	42	31	30	86	4.8	4.0	9.0	4.9	3.1	2.2
Grade VI . . .	50	38	37	99	5.8	5.0	11.0	6.6	3.7	2.8
Grade VII . . .	58	44	44	110	6.8	6.0	13.0	8.3	4.2	3.4
Grade VIII . .	63	49	49	117	7.8	7.0	14.0	10.0	4.8	4.0
Grade IX . . .	65	50	50	120	8.6	7.8	15.0	11.0	5.0	4.3
Time allowances, minutes . . .	1	1	1	1	6	6	12	12	6	6

Thus in the Addition Test (Test No. 1), the average score in Grade V is 42, the number of correct additions made in one minute. Similarly, for all the other tests.

## II. WOODY ARITHMETIC SCALE<sup>1</sup>

Whereas, in each of the separate Courtis Tests the problems are of approximately the same difficulty throughout, in the Woody Scales a different method of measuring efficiency is employed. The scales are designed to measure work in the four fundamental operations of (a) addition, (b) subtraction, (c) multiplication, and (d) division, respectively. Each of these scales consists of a great variety of problems falling within the field of the particular operation that the scale is designed to test. These problems, beginning with the easiest that can be found, gradually increase in difficulty until the last ones in each scale are so difficult that only a relatively small percentage of the pupils in the eighth grade are able to solve them correctly. That is, taking the addition scale for example, the problems rise in difficulty from the first, which requires next to no ability in addition, up to the last, which, though still an addition problem, is of sufficient complexity to test children of the eighth grade. The relative difficulties of the problems within each scale were determined by administering them to large groups of children in several school systems, the difficulty of a problem being calculated from the percentage of correct answers by a method similar to that used in the Buckingham Spelling Scale.

Two distinct series of scales in each of the above named operations have been devised. It will be sufficient here to describe the shorter of these scales, Series B, and to illustrate the general principles which underlie this method of measurement. For the other scale with a full account of its instructions, method of administration, scoring, etc., the reader is referred to the original study.

<sup>1</sup> The scales are reproduced by the courtesy of Dr. Clifford Woody.

## Series B — Addition Scale

Name.....

When is your next birthday?.....How old will you be?.....

Are you a boy or girl?.....In what grade are you?.....

(1)	(2)	(3)	(5)	(7)	(10)
2	2	17	72	$3 + 1 =$	21
3	4	2	26		33
—	3	—	—		35
	—				—

(13)	(14)	(16)	(19)	(20)
23	$25 + 42 =$	9	\$ .75	\$12.50
25		24	1.25	16.75
16		12	.49	15.75
—		15	—	—
		19		
		—		

(21)	(22)	(23)	(24)	(30)
\$8.00	547	$\frac{1}{2} + \frac{1}{2} =$	4.0125	$2\frac{1}{2}$
5.75	197		1.5907	$6\frac{3}{8}$
2.33	685		4.10	$3\frac{3}{4}$
4.16	678		8.673	—
.94	456		—	
6.32	393			
—	525			
	240			
	152			
	—			

(33)	(36)	(38)
.49	2 yr. 5 mo.	$25.091 + 100.4 + 25 + 98.28 + 19.3614 =$
.28	3 yr. 6 mo.	
.63	4 yr. 9 mo.	
.95	5 yr. 2 mo.	
1.69	6 yr. 7 mo.	
.22	—	
.33		
.36		
1.01		
.56		
.88		
.75		
.56		
1.10		
.18		
.56		
—		

Series B — Subtraction Scale

Name.....  
 When is your next birthday?.....How old will you be?....  
 Are you a boy or girl?.....In what grade are you?.....

(1) 8 5 —	(3) 2 1 —	(6) 11 7 —	(7) 13 8 —
(9) 78 37 —	(13) 16 9 —	(14) 50 25 —	(17) 393 178 —
(19) 567482 106493 —	(20) $2\frac{3}{4} - 1 =$	(24) $8\frac{7}{8}$ $5\frac{1}{2}$ —	(25) 27 $12\frac{5}{8}$ —

(27) 5 yds. 1 ft. 4 in. 2 yds. 2 ft. 8 in.	(31) $7.3 - 3.00081 =$	(35) $3\frac{1}{8} - 1\frac{5}{8} =$
--	---------------------------	---

Series B — Multiplication Scale

Name.....  
 When is your next birthday?.....How old will you be?.....  
 Are you a boy or girl?.....In what grade are you?.....

(1) $3 \times 7 =$	(3) $2 \times 8 =$	(4) $4 \times 8 =$	(5) 23 3 —
(8) 50 3 —	(9) 254 6 —	(11) 1036 8 —	(12) 5096 6 —
(13) 8754 8 —	(16) 7898 9 —	(18) 24 234 —	(20) 287 .05 —
(24) 16 $2\frac{1}{4}$ —	(26) 9742 59 —	(27) 6.25 3.2 —	(29) $\frac{1}{2} \times 2 =$
(33) $2\frac{1}{2} \times 3\frac{1}{2} =$	(35) $987\frac{3}{4}$ 25 —	(37) $2\frac{1}{2} \times 4\frac{1}{2} \times 1\frac{1}{2} =$	(38) $.0963\frac{1}{2}$ .084 —

## Series B — Division Scale

Name.....

When is your next birthday?.....How old will you be?.....

Are you a boy or girl?.....In what grade are you?.....

(1)	(2)	(7)	(8)
$3\overline{)6}$	$9\overline{)27}$	$4 \div 2 =$	$9\overline{)0}$
(11)	(14)	(15)	(17)
$2\overline{)13}$	$8\overline{)5856}$	$\frac{1}{2}$ of 128 =	$50 \div 7 =$
(19)	(23)	(27)	(28)
$248 \div 7 =$	$23\overline{)469}$	$\frac{7}{8}$ of 624 =	$.003\overline{)0936}$
(30)	(34)	(36)	
$\frac{3}{4} \div 5 =$	$62.50 \div 1\frac{1}{4} =$	$9\overline{)69}$ lbs. 9 oz.	

Series B was especially constructed for use in the measurement of arithmetical ability when the amount of time for such measurement is limited. The break in continuity in the numbering of the problems does not mean that the whole scale is not presented. The scale is quite complete as it stands; the numbering is a matter of convenience for purposes external to the use of the scale.

The Addition and Subtraction Scales can be used in Grades II to VIII inclusive; the Multiplication and Division Scales, in Grades III to VIII inclusive. It is recommended that in the use of Series B all tests be given together.

## DIRECTIONS FOR ADMINISTRATION

It is very necessary that the same standard method be employed in the giving of these tests; care should be taken that the same directions are given in the same way to all groups taking the tests. The following are the general directions which should be carefully followed: Distribute the papers face down and do not allow the pupils to turn them over until they are told to do so. When all are ready with pencils in hand, say: "Turn your papers over and answer the questions at the top



of the page." When all these preliminary questions have been answered, repeat the following formula of specific directions. If you are giving the Addition test, say, "Every problem on the sheet which I have given you is an addition problem, an 'and problem.' Work as many of these problems as you can and be sure that you get them right. Do all your work on this sheet of paper and don't ask anybody any questions. Begin."

For each test in Series B allow ten minutes. It is essential that all the pupils start and stop work together because the test is partly one of speed. Most of the children will have finished all that are within range of their ability before the end of the time allowed; those who have not must not be allowed any further time.

The only variation in procedure in giving any of the other tests is the substitution in the formula of specific directions of the expressions "subtraction or 'take away problems,'" "multiplication or 'times problems,'" and "division or 'into problems,'" for the expression "addition or 'and problems.'" Since teachers in the lower grades sometimes use the expressions "and," "take away," "times," and "into," problems, these forms should also be used in administering the test so as to make clear to the children what is expected of them.

#### DIRECTIONS FOR SCORING THE TESTS

In scoring each test the standard of marking should be absolute accuracy and the final answer should be in its lowest terms.

If the results of class measurement are to be compared with the results and values established by the author, only those answers should be accepted as correct which are identical with those given in the following table, since these are the solutions upon the basis of which the original scoring was done.

## ANSWERS TO PROBLEMS. SERIES B

ADDITION		SUBTRACTION	
Problem	Answer	Problem	Answer
1 . . . . .	5	1 . . . . .	3
2 . . . . .	9	3 . . . . .	1
3 . . . . .	19	6 . . . . .	4
5 . . . . .	98	7 . . . . .	5
7 . . . . .	4	9 . . . . .	41
10 . . . . .	89	13 . . . . .	7
13 . . . . .	64	14 . . . . .	25
14 . . . . .	67	17 . . . . .	215
16 . . . . .	79	19 . . . . .	460,989
19 . . . . .	\$2.49	20 . . . . .	$1\frac{3}{4}$
20 . . . . .	\$45.00	24 . . . . .	$3\frac{1}{8}$
21 . . . . .	\$27.50	25 . . . . .	$14\frac{3}{8}$
22 . . . . .	3,873	27 . . . . .	2 yds. 1 ft. 8 in. not 8 lin.
23 . . . . .	$1\frac{2}{3}$	31 . . . . .	4.29919
24 . . . . .	18.3762	35 . . . . .	$2\frac{1}{4}$ not $2\frac{2}{3} = \frac{1}{4}$
30 . . . . .	$12\frac{5}{8}$ not $11\frac{1}{8} = 1\frac{1}{8}$		
33 . . . . .	10.55		
36 . . . . .	22 yrs. 5 mo. or $22\frac{5}{12}$ yrs.		
38 . . . . .	268.1324		

MULTIPLICATION		DIVISION	
Problem	Answer	Problem	Answer
1 . . . . .	21	1 . . . . .	2
3 . . . . .	6	2 . . . . .	3
4 . . . . .	32	7 . . . . .	2
5 . . . . .	69	8 . . . . .	0
8 . . . . .	150	11 . . . . .	$6\frac{1}{2}$ not $6 + 1$
9 . . . . .	1,524	14 . . . . .	732
11 . . . . .	8,288	15 . . . . .	32
12 . . . . .	30,576	17 . . . . .	$7\frac{1}{2}$ not $7 + 1$
13 . . . . .	70,032	19 . . . . .	$35\frac{2}{3}$ not $35 + 3$
16 . . . . .	71,082	23 . . . . .	$20\frac{2}{3}$ ; 20.3, not $20 + 9$
18 . . . . .	5,616	27 . . . . .	546
20 . . . . .	14.35	28 . . . . .	31.2
24 . . . . .	42	30 . . . . .	$\frac{3}{20}$ or .15
26 . . . . .	574,778	34 . . . . .	50
27 . . . . .	20,000	36 . . . . .	71 lbs. $11\frac{2}{3}$ oz.; 71 lbs. 9 oz.
29 . . . . .	$\frac{1}{4}$ not $\frac{2}{4}$		
33 . . . . .	$8\frac{1}{4}$		
35 . . . . .	$24693\frac{3}{4}$		
37 . . . . .	$15\frac{1}{8}$		
38 . . . . .	.0080902 $\frac{1}{2}$ or .00809025		

## METHOD OF DETERMINING THE CLASS ACHIEVEMENT

The method used for determining the class achievement with Scale B is simpler than that employed in the use of Scale A. It is largely for this reason that Scale B was chosen for description. It should be noted that in each of the scales a definite attempt has been made to place the problems so that they would increase by uniform stages of difficulty from the first to the last. Thus, in the Addition Scale problem 3 is as much more difficult than problem 2 as problem 5 is more difficult than problem 3, and so on. If one compares this with the method of the Courtis Tests, it will be seen that in the latter the problems involving a given operation are all of approximately the same difficulty and require precisely the same knowledge and method for their solution. In other words the Courtis Tests measure speed in the various operations in arithmetic rather than extent of knowledge of the operation involved. In the Woody Scale, because the problems increase in difficulty, the score measures a certain extent of knowledge of the process involved in the operation rather than mere speed of performance. For example, in a race one could have a series of hurdles of all the same height and test the number cleared in a certain time — such is the Courtis method; or the hurdles could get gradually higher and higher, the success of the individual being measured by the hurdles he can clear without a fall — such is the method of the Woody Scales.

An objection is sometimes made by teachers that the problems are too hard for the children. In this connection it cannot be pointed out too clearly that when scales of this type are used in the schools it is not expected that the children will be able to do all the problems, just as when we determine the height of a child by means of an eight-foot rule, we do not expect the child to measure up to the eight feet.

The achievement of a class is measured by calculating the median number of problems which were solved correctly. By the median number is meant that number which marks the point at which there are just as many pupils who solve a greater number correctly as there are those who solve a less number correctly. In order to measure the median point of achievement of the class, it is necessary to make a distribution table, showing the number of pupils who were unable to solve a single problem correctly, the number who solved one, two, three, etc., up to the final number. Take the following as an example:

NUMBER OF TIMES A GIVEN NUMBER OF ADDITION PROBLEMS WAS  
SOLVED CORRECTLY

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
No. of pupils . . . .	1	0	2	3	0	4	1	2	3	5	6	11	1	5	2	4	1	1	

That is, one pupil failed to solve a single problem. With that exception there were no children who did not solve at least one problem correctly. Two children solved two problems correctly, three children solved three problems correctly, four solved five correctly, and so on.

Since there are, let us say, 52 individuals in a given class, "the median point evidently falls between the achievements of the 26th and 27th pupils. Let us begin with the individual who was unable to solve a single problem correctly and count the two individuals who solved two problems, the three who solved three problems, and so on until we come to the step that includes the 26th individual. Now if we are to indicate the exact point in the achievement of the pupils where there are just as many pupils who solve a greater number of problems as there are those who solve a less number, it is necessary to count 5 of the 6 individuals who solved 10

problems correctly. Thus, on the assumption that the individuals are distributed over any step at equal distances from one another, the median point is  $\frac{5}{8}$  of the distance through this step. Hence, the median achievement of this class, *i.e.* the median number of problems solved, is 10.8 problems correctly solved."

### TENTATIVE STANDARDS OF ACHIEVEMENT

The following standards of achievement have been determined on the basis of tests made on several thousand children from the second to the eighth grades of various school systems. It is possible that with further experimentation they may need to be slightly altered.

TENTATIVE STANDARD OF ACHIEVEMENT FOR SERIES B

GRADE	ADDITION	SUBTRACTION	MULTIPLICATION	DIVISION
II . . . .	4.5	3		
III . . . .	9	6	3.5	3
IV . . . .	11	8	7	5
V . . . .	14	10	11	7
VI . . . .	16	12	15	10
VII . . . .	18	13	17	13
VIII . . . .	18.5	14.5	18	14

The standards are based upon the total number of problems that were correctly solved in each grade. Thus in the second grade in addition, the median achievement was 4.5 problems, in the third grade, 9.0 problems, etc.

All that is necessary, therefore, to test a class is to procure the standard blanks, follow the detailed instructions in administration and scoring, and then determine the median score by the method shown. This median score can then be compared with the tentative scores given by the author. It should be noted of course that these tentative scores would cease to have significance if, previous to the test, the children had been drilled on examples framed with the particular scale problems in mind.

## EXERCISES

1. How does the general character of the work of your class, as revealed by the administration of Test No. 7, Series A, compare with that of other classes of the same grade in your building or city?
2. How does the work of your pupils in the various sub-branches, as revealed by the tests, compare with the standard scores for your grade? How does it compare with the work in these sub-branches in other classes and schools where you may be able to test?
3. Suppose the frequent administration of the tests failed to reveal a reasonable amount of improvement in the various sub-branches, what would this seem to indicate?
4. Could the tests be utilized to remedy this condition?
5. What two important facts in regard to the ability of the pupils in your class have the tests revealed?
6. Suppose the tests showed the ability of a pupil to differ greatly in the various sub-branches, what action should the teacher take in regard to it?
7. For what purposes may the Woody Scale be used to greater advantage than the Courtis Scale?
8. In your experience with the tests, have they tended to show any relation between ability in one branch and ability in another?
9. What precautions should be taken in administering the tests?
10. In what ways should the continued use of these tests increase the efficiency of a teacher of arithmetic?

## CHAPTER III

### HANDWRITING SCALES

#### I. THORNDIKE SCALE

#### II. AYRES SCALE

#### III. COURTIS TESTS

PROBABLY there is no subject about which opinions of efficiency are more vaguely expressed than the subject of handwriting. Such terms as "good," "fair," "poor," etc., merely express the individual teacher's judgment as determined by certain factors, such as legibility, grace, character, etc., or by certain styles, such as vertical or slanting, to which that individual is partial. No two teachers mean the same quality by the use of the same term. Consequently such judgments, because they are not expressed in terms of a universal standard which conveys the same meaning to everybody, are of little value when comparisons are necessary. Within recent years attempts have been made to eliminate this unscientific type of judgment, which is the natural result of the lack of a standard, by the construction of a scale for measuring the quality of handwriting. Thorndike and Ayres have each devised such a scale or standard, while Courtis has outlined a method by which it is possible to obtain samples of children's handwriting, made under uniform conditions. Each of these methods will be described briefly in turn.

THE THORNDIKE SCALE FOR HANDWRITING

4 started on the  
carriage my

32

5 bus her and the carriage  
arrived along down the  
driveway. see and see



6 gathering about them and  
took away in an instant leaving  
only a poor old lady

7 and, John vanished behind the  
bushes and the carriage moved

8 moved along down the driveway. The audience of passers-by, which had been gathering about them melted away.

Then the carelessly gentleman step -  
ped lightly into Warren's carriage and  
held out a small card, then vanished be-  
hind the bushes and the carriage moved

Then the carelessly dressed gentlemen stepped lightly into Warren's carriage and held out a small card, John Vanished behind the by which had been gathering about them melted away in an instant leaving only a poor old lady on the curb. Albert was sadly

Then the carelessly dressed gentleman stepped lightly into Warren's carriage more and held out a small card, John vanished

**10** driveway. The audience of passers-by, which  
had been gathering about them melted away  
in an instant leaving only a poor old lady on  
the curb. Albert was sadly stricken.

11 rage moved along - down the driveway. The audience of passers-by which had been gathering - about them melted away

along ~~the~~ down the driveway. The audience of passers-by which had been gathering about them

John vanished behind the bushes and the carriage moved along down the driveway. The audience

12 lightly into Warren's carriage and held out a small card, John vanished behind the bushes and the carriage moved along down the driveway behind the bushes and the carriage moved along down the driveway. The audience of passers-by

Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small card, John vanished behind the bushes and the

Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a shield behind the bushes and the carriage moved along down the driveway. The audience of passers-by which had

Then the carelessly dressed gentleman stepped lightly into Warren's carriage and

Then the carelessly dressed gentlemen stepped lightly into Warren's carriage and

14 Then the carelessly dressed gentleman  
stepped lightly into Warren's carriage  
and held out a small card, I

Then the carelessly dressed gentleman  
stepped lightly into Warren's carriage and  
held out a small card, John vanished behind



15 lightly into Warren's carriage and held out a small card, John vanished behind the bushes and the carriage moved along down the drive

*held out a small card, John vanished behind the bushes and the carriage moved along down the driveway. The audience of passers-*

John vanished behind the bushes and the carriage moved along down the driveway. The audience

Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small white

Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small

gentleman stepped lightly into Warren's carriage and held out a small card, John vanished.

17 Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small card, John vanished be-

18 showed that the rise and fall of the tides the attraction of the moon and sun upon

The foregoing scales are reproduced by the courtesy of Dr. E. L. Thorndike.

## I. THORNDIKE HANDWRITING SCALES

Thorndike was the first to construct an objective scale for handwriting. This appeared March, 1910, and was developed as follows. One thousand samples of handwriting, ranging from the worst to the best to be found in the sixth, seventh, and eighth grades, were given in turn to forty competent judges. Each of these judges was asked to rank these samples according to their "general merit," which was to be based on a combination of grace and legibility, by placing each specimen in one of eleven arbitrary groups in order of increasing merit. Previous experiments had shown that these samples, instead of falling into a thousand different classes, naturally fell into about eleven groups, all the members of a group being of about equal merit. That is, the same thing is true of handwriting as is true of attempting to divide into a thousand classes a thousand people whose height varies from five to six feet. Many would be so nearly of the same height as to make such a classification impracticable, if not impossible. Similarly, exact classification would be impossible in the case of writing, where the distinction between the samples was not pronounced.

After each judge had placed each sample three or four times in this way in one of these eleven groups, the average result of his rankings was taken as his final grading for each specimen; that is, if a judge ranked a certain specimen of handwriting in class 10 on the first occasion, in class 11 on the second, in class 12 on the third and in class 10 on the fourth, on the whole he placed it somewhere between classes 10 and 11, or to be exact, at a point which can be represented by 10.7. Then the returns of all the judges were massed and the average of all rankings given to each sample was determined. In this way the place assigned to each specimen by the combined opinion of all the judges was fixed. When the averaged judgments were collected (as might be expected

where so many samples were concerned), it was found that some samples were placed in, or approximately in, each of the eleven groups; that is, some samples were graded 1, 2, 3, 4 . . . 11, while many samples were given rankings midway between the different groups, indicated by the markings 1.4, 1.6, 2.1, 2.8, etc.

Now when it is recalled that each one of these groups, in the opinion of the judges, is separated from the others by equal steps of merit, it may readily be seen how a handwriting scale can be obtained, provided only that samples be graded exactly or approximately as falling into groups 1, 2, 3, 4, . . . 11, the handwriting samples in group 2 being as much superior to those in group 1 as those in group 3 are superior to those in group 2, etc. In this way the Thorndike Scale was obtained, a scale whose steps of difference forty competent judges have considered to be equal. Later, this scale was extended to include fifteen classes of handwriting which ranged in quality from handwriting which may barely be called such to that suitable for decorative purposes.

This scale with its various classes of handwriting has from one to three different styles of writing in each group. Undoubtedly it would be far more satisfactory if each class contained samples of all the various types of writing which are found in the school. This defect, however, can easily be remedied when a larger number and greater variety of samples become available. Furthermore, it is to be regretted that this scale, which measures about twenty-two by twenty-four inches, is not issued in more convenient form.

In spite of these slight defects, which time will remedy, the scale is certainly far superior to the judgment of any one individual. The method of using it is very simple. A sample of handwriting is measured by placing it alongside the scale and estimating to which one of the fifteen groups, as represented by the fifteen samples, it belongs.

If it is thought to lie between two groups, a fraction may be added or subtracted according to whether it is judged better or worse than the sample on the scale to which it most nearly corresponds. Thus, if it falls between classes 12 and 13, it might be graded at any point in between, such as 12.4 or 12.8. For especially accurate work, it is well to have several individuals rank the samples of handwriting and then take the average of their rankings as the final measurement. Care should be taken to decide a specimen's grade not because of its likeness in *style* to some sample in the scale, but because of its likeness in *quality*.

After the person grading has become familiar with the scale, comparisons will be facilitated if the scale is folded so that the samples form the pages of a book. Then the judge should pass rapidly from the lowest to the highest sample, rating the specimen by his impression as a whole, inasmuch as such an impression is the resultant effect of all the qualities possessed by the writing. Long, painstaking comparisons prevent accuracy instead of securing it. When it is necessary to compare a specimen with samples unlike it in slant and character, placing it somewhere between two groups will often solve the difficulty.

## II. AYRES HANDWRITING SCALE<sup>1</sup>

The Thorndike Scale is based on *general merit* of handwriting. The Ayres Scale, on the other hand, is based on *legibility*; thus there is a substitution of function, instead of appearance, as a criterion. Ayres takes this standard for two reasons. In the first place, the purpose of writing is to be read; hence "readability," or legibility, is the prime requisite. In the second place, it is exceedingly easy to measure the legibility of any sample of handwriting by determining the time it takes to read it. In this

<sup>1</sup> For reproduction of Ayres Scale, see pages 50 to 57.

way an exact evaluation of the relative legibility of any specimens may be obtained in terms of a unit of time. The criterion of general merit, though based on the opinion of competent judges, does not allow of this accuracy.

The method by which this scale was produced differs radically from that used by Thorndike. Previous experiments had shown that the best way to find out the relative legibility of different samples of handwriting was to find out the rate, in words per minute, at which each sample could be read. In order to represent a random selection and not the writing of any particular city or section, 1578 samples were secured of the handwriting of children in the upper elementary grades of 40 school systems in 38 different states. These samples did not consist of words so arranged as to convey a meaning, but were composed of words thrown out of context. The object of this was to make it necessary for the reader to decipher each word separately, and to make it impossible for him to memorize. Through the coöperation of superintendents and teachers, samples from either the best or the worst class in any city were avoided, and it was so arranged that the pupils made no effort to write with exceptional care or rapidity.

These 1578 samples were then turned over to ten competent paid assistants, who in turn read each sample and by means of a stop watch recorded the exact time it took to read it. After each sample had been read by the ten readers, the average time taken to read it was computed. Then the rate in words per minute at which the reading had been accomplished, was found by dividing the average time it took to read a given sample by the number of words in it. This process was repeated for each one of the 1578 samples. After it had been determined to what extent the readers had increased their reading speed through practice, the first 75 papers were reread and new times recorded to correct this error.

The next step was the classification of these samples. After various attempts at this, five classes — vertical, medium slant, extreme slant, backhand, and mixed — were finally carefully defined on the basis of the arbitrary judgment of a number of competent judges. Then each of the samples was classified on the basis of the slant of its letters and assigned to one of these five classes. Because of the limited number of backhand and mixed samples of handwriting, these were left out of the final scale.

The scale itself was then constructed in the following manner. All of the samples, which had been so marked as to indicate both the rate at which each one had been read and the class or style of writing — vertical or slant, etc. — to which it belonged, were arranged in one long series beginning with the sample having the lowest time rating and extending to the one having the highest. As might be expected, there were many samples of medium grade — that is, that were read at a medium rate; only a few that were very good — read at a rapid rate; and only a few that were poor — read at a very slow rate. Then, beginning at the poorest sample — that which took the longest time to read — a count was made just halfway through the samples. The specimen thus obtained was the central point, below which one half of the samples were read more slowly, and above which, one half were read more rapidly. This sample had been marked 175.7 indicating that it was read at the rate of 175.7 words per minute. Because of its central position, considering the entire series as 100, this sample was called 50. In a similar manner samples were picked out one tenth, two tenths, three tenths, four tenths, six tenths, seven tenths, eight tenths, and nine tenths of the way through the series, and these were designated 10, 20, 30, 40, 60, 70, 80 and 90, respectively. These values were chosen because teachers are familiar with them in grading.



The rate of reading marked on these samples was found to be 130.2, 149.4, 163.5, 175.7, 186.1, 195.8, 202.9, and 209.6 words per minute, respectively. Thus it was seen that this scale does not proceed by equal steps as far as the time consumed in reading is concerned. Instead, the gain in time rate became progressively smaller as one moved from the worst to the best sample. How reasonable this is, may easily be seen. A very poor handwriting takes a long time to decipher. One which is just slightly better may be read almost twice as fast. A still better one may be read somewhat faster but not twice as fast as the previous one, and so on, the gain in the rate growing smaller and smaller as the handwriting improves. Thus, as far as readability is concerned, the difference in the time it takes to read a sample marked 30 and one marked 40 is greater than the difference in the time it takes to read one marked 60 and one marked 70. So, what is actually meant when it is said that the steps of this scale are equal, is that each one of them has been so chosen that it is as much better than the one before, as that one is better than the preceding one. Qualities 60 and 40 are respectively equally distant above and below quality 50; that is, there is the same proportion of samples between 50 and 60 as between 40 and 50, and so on down the scale.

The scale itself is on a sheet of paper measuring nine by thirty-six inches. It contains eight groups (the lowest and the highest groups being omitted in the final scale), each group including three types of handwriting, the vertical, the slant, and the extreme slant. Ayres' studies have shown that 95% of the common writings of school children are included in these three styles. To facilitate comparison, both the paper and the ink used in the scale are of the color used in the public schools. The scale is used in exactly the same way as the Thorndike Scale.

The following scales are reproduced by the courtesy of Dr. Leonard P. Ayres.

Beard, his trusty fowlin  
and army of name him  
he who demanded that he  
be well deservment his of  
Confunded completely you was

The great error in Rip's composition w  
kind of profitable labor it could not bly  
condemned worse the worst yet potatoes as  
left more little was there until aced by a  
had estate Patrimonial his through that a

parlor the painting, &  
figures the of Rip & his  
then in cross with shot  
stocking old feather and  
hanger and belt & pouch

brated have I as much  
all cream and milk of bou  
roasted and shad broiled  
pears and peaches and plurr  
then and cakes sweet cru

The hair of the affright  
upon his head with  
the clove tongue his  
his resume to endea  
To turn and turn he

It was toward evening that he  
friend with he found thou  
of unworthy as horse brake he  
reck his of risk constant ai  
of kinds all to given amusin

The appearance of Rys, beard, his rusty fowl in and the army of name I who demanded hee eve wilderness of his misde

The hair of the offrighted pedu terror What was to be done? him within sink to begs other his quickened however leaving of hopes in steel hi

The gallant hero now spe hour at his toilet brush his best country the in, looked he as down bro had and rider furious a ci

Fain would I pause to dwell  
 burst upon the enraptured g  
 every to justice ample did I  
 great so in not was here ou  
 on get to eager too am ar

His school was a low building  
 constructed of logs the window  
 partly strong the of those of  
 the of back off burden the ta  
 studying pupils his of mu

Schabod pride himself  
 much as upon his vocal  
 not a fibre about exple  
 himself make to recol  
 and fiction becoming,

He entered the house, which  
 had always kept in the  
 general character large  
 and hat cocked a with a  
 sceptre or of instead has

It is remarkable that the  
 I have mentioned is not  
 inhabitants of the valley  
 vegetation families same  
 the find still not should

Ishabod pride himself  
 much as upon his a  
 limb, not a fibre about  
 her the himself me  
 his of little the in

his by called city gre  
 the upon eye gardia  
 terprise his of scenes  
 this in permitted be  
 been always had r

At length he reached to w  
 opened through the cliffs  
 atre but no traces of lov  
 beard his found he astor  
 same the do to involu

stranger's appearance  
 square built familia  
 awe inspired that  
 about incompreh  
 bright the and s

The appearance of Rip, with his  
rusty fowling piece, his uncon-  
of name his was what ar-  
ed hat cocked the in man  
midst the in. Man another

As Chabod jogged  
way his eye, ever o-  
tom of culinary ar-  
with delight Hudso  
the in sounds an

On nearer approach  
surprised at the se-  
stranger's appearan-  
square built form  
and awe inspired



We had not been home long w  
 was heard from the distance A  
 good before way gave soon gu  
 Daughters their with romped  
 the with gossiped cottage and

We had not been home  
 of music was heard f  
 of country cheer good be  
 guests the of bashful  
 not heard been music

Your mere puny stri  
 at the flourish of t  
 passed by with in  
 claims hankerc  
 their at more seer

### III. COURTIS HANDWRITING TESTS

It is apparent that the efficient use of such objective scales necessitates imposing uniform conditions when obtaining the handwriting samples to be measured. That is, it would be unfair to grade by the same standard a sample of handwriting written at a very rapid rate of speed, and one written at a very slow rate of speed. The time allowed in the writing of the samples should be the same. Curtis has attempted to overcome this difficulty by the use of a simple method whereby samples may be obtained of a pupil's handwriting under formal conditions (at a fixed rate of speed), and under casual conditions (at the writer's natural rate of speed). Samples written under either one of these conditions, providing all the samples were obtained under the same circumstances, may then be graded by either the Thorndike or the Ayres scale.

The value of such scales as have been described is very great. In the first place, they represent measures that are the outcome of the thought, labor, and experiment of many persons and as such they are far superior to the casual and oftentimes thoughtless opinion of any one individual. Not only may such scales measure the relative quality of different samples of handwriting or the improvement in handwriting in an individual, but by means of them entire classes, groups, and systems, whether chosen for grade, age, sex, method of teaching, length of practice drills, etc., may be compared; for the scale constitutes the basis of judgment wherever it is used. So it makes no difference whether a sample is called *X*, *A*, 100, or any arbitrary name, provided it represents in terms of the scale the same quality to every person. By means of these scales teachers may determine the progress and needs of their pupils as well as the efficiency of the methods they are using. Supervision may also be made

more effective because school officials, given the actual results of such tests, are provided with the means for making a scientific analysis of school conditions. However, as Thorndike says, such scales will do their greatest service not as measuring rules, but "by creating in the minds of teachers a mental standard to be used in even the most casual ratings of everyday school life."

### STANDARD SCORES

It is obvious that if the above scales are used in measuring the writing ability of a sufficient number of children, standard scores of efficiency for each grade may be computed by finding the arithmetical average of the records obtained in each grade. That is, on the basis of the results obtained from the use of the scales, it is possible to say just how fast and how well pupils in any one grade should write.

Starch secured tentative standard scores of this kind by administering the Thorndike Test at the end of the school year to about 4000 pupils in eight cities of three states, and finding (1) the average rate of speed and (2) the average quality of writing, in each one of the grades. Starch's results were as follows:

Grades . . . . .	1	2	3	4	5	6	7	8
Speed (letters per minute)	20	31	38	47	57	65	75	83
Quality (Thorndike Scale)	6.5	7.5	8.2	8.7	9.3	9.8	10.4	10.9

According to these scores we find that the average child in the second grade, for instance, writes at a speed of 31 words per minute and possesses a quality of handwriting indicated by 7.5 on the Thorndike Scale or, by derivation, 26.5 on the Ayres Scale.

The Thorndike Scale was chosen in preference to the Ayres Scale because the latter does not extend so far at the lower and upper limits as the former; that is, the

limits of the Ayres Scale lie within qualities 7 to 14 of the Thorndike Scale. However, equivalent values for qualities in the Ayres Scale have been derived; that is, one step on the Thorndike Scale has been found to be equal to 8.9 points on the Ayres Scale. So the standard scores may be expressed in units of either scale. As a result of this derivation we have the following equivalents:

THORNDIKE SCALE			AYRES SCALE	
Quality	7	is equal to	quality	22
"	8	" " "	"	31
"	9	" " "	"	40
"	10	" " "	"	49
"	11	" " "	"	58
"	12	" " "	"	67
"	13	" " "	"	76
"	14	" " "	"	85

In obtaining these standards it was apparent (1) that writing in the various grades differed widely in regard to both speed and quality; and (2) that the abilities of the pupils in one grade in many instances overlapped the abilities of the pupils in the adjacent grades. For instance, the handwriting of pupils in the first grades in three schools in city A was found to range all the way from quality 4 to quality 10.5, and that of pupils in the eighth grade, from quality 5 to quality 15. So great was the overlapping that the averages of the various grades differed from each other by only small amounts.

To recapitulate, the first thing to be done by a teacher who is desirous of measuring the handwriting of pupils, is to see that the samples to be measured are all obtained under the same conditions. This may be done by having the pupils write at their natural rate of speed a selection, or part of a selection, with which they are all familiar. Care should be taken to see that they all start writing at the same time and stop at the same time, say at the end of two minutes.

The speed of writing for each pupil may be easily determined, in terms of letters per minute, by dividing the total number of letters written by two. The average speed of writing for the class may then be computed and compared with the standard score for that grade as given by Starch.

The samples may then be measured for their quality by either the Thorndike or the Ayres Scale. The former may be secured by sending to Teachers College, Columbia University, New York; and the latter, by sending to the Russell Sage Foundation, Department of Education, New York.

In determining the quality of any given specimen of handwriting, all that is necessary is to slide the specimen along the scale beginning with the poorest sample until a writing of corresponding quality is found. If this writing is marked, say 11 in the Thorndike Scale, then that is the value of the handwriting measured. If the quality of the specimen seems to be better than quality 11, but not so good as quality 12, then it should be given a value somewhere between 11 and 12, say 11.2 or 11.8, according to whether it is judged better or worse than the sample on the scale to which it most closely corresponds.

It is advisable for the teacher to have one of the scales on the wall, in order that it may be utilized by the pupils themselves.

It is obvious that in comparing the handwriting of any two classes of the same grade the following conditions must have been fulfilled: (1) All of the children must have written the same selection for their samples; (2) all must have had the same degree of familiarity with the selection; (3) all must have been allowed the same time in which to write; (4) all the results must be expressed in terms of the same scale. Such uniform conditions may be realized within any given school system through a set of specific instructions sent out by the supervisor of

writing or the superintendent of instruction some days before the measurements are to be made.

### EXERCISES

1. Take thirty specimens of writing distributed through the grades and have these marked by five teachers according to their usual percentage methods. How do the results agree?

2. Repeat the above experiment with the exception that the grading is done by means of the Thorndike Scale. How do the estimates of the five teachers compare now?

3. Suppose a teacher found as a result of using the scale that the pupils were above the average (standard score) for that grade in speed of writing but below the average in quality, or vice versa, what should be done?

4. Does there seem to be any relation between speed and quality of handwriting?

5. Is there any practical lesson to be learned from the fact that, although a group of children may be made to write in the same way up to 14 years of age, at 18 each has his own particular style?

6. What would be the result of getting the children to grade their own handwriting and to compare their results?

7. Suppose a teacher found a great difference in the quality of the handwriting of the pupils in a given grade, some pupils writing far above the average for that grade and others far below it, what should be done?

8. What is the best thing to do if a teacher finds it necessary to compare a given specimen of handwriting with samples in the scale which are unlike it in slant or character, or both?

9. How may the scale be used to test the efficiency of any method of teaching handwriting?

10. What factors should a teacher take into consideration when setting a standard of handwriting for a given class?

## CHAPTER IV

### READING SCALES

#### I. THORNDIKE AND GRAY SCALES

#### II. STARCH SCALE

#### III. COURTIS SCALE

SINCE it is through reading that a large part of our information is obtained, an objective means of measuring efficiency in that subject is of great importance. Several attempts have been made to fill this need. In 1914 Thorndike and Gray published tentative scales for measuring school achievement in reading, and later both Starch and Courtis published tests for the same purpose.

#### I. THORNDIKE AND GRAY SCALES

In Thorndike and Gray's Scales attempts are made to measure the following factors: "(1) silent reading so far as it concerns (*a*) the understanding of words singly and (*b*) the understanding of sentences and paragraphs; and (2) simple, oral reading of matter-of-fact passages." Each of these three scales will be described in turn.

##### (1) Scale A for Visual Vocabulary

The scale designed to measure a pupil's knowledge of the meaning of single words is called "Scale A or the Scale for Extent or Range of Visual Vocabulary." It is printed on a single sheet of paper and consists of nine lines of words. These lines are numbered 4, 5, 6, 7, 8, 9, 10, 10.5 and 11 respectively. All the words on the same line are about equally hard to understand, and their difficulty increases gradually from line to line. The first

line is marked 4 because the difference between a child who can read the first line and one who can read nothing at all is about four times as great — measured in years of work — as the difference between a child who can read line 4 and one who can read line 5, or as the difference between pupils who can read any other two successive lines. The seventh line is marked 10.5 because the words on it are a little too hard to stand half-way between lines 9 and 11. There are only three words on line 11 because no others of precisely the same difficulty could be found.

### Thorndike Reading Scale A

#### Visual Vocabulary

Write your name here . . . . .

Look at each word and write the letter F under every word that means a *flower*.

Then look at each word again and write the letter A under every word that means an *animal*.

Then look at each word again and write the letter N under every word that means a *boy's name*.

Then look at each word again and write the letter G under each word that means a *game*.

Then look at each word again and write the letter B under every word that means a *book*.

Then look at each word again and write the letter T under every word like *now* or *then* that means something to do with *time*.

Then look at each word again and write the word GOOD under every word that means something *good to be or do*.



---

Then look at each word again and write the word  
BAD under every word that means something  
*bad to be or do*.

4. camel, samuel, kind, lily, cruel
5. cowardly, dominoes, kangaroo, pansy, tennis
6. during, generous, later, modest, rhinoceros
7. claude, courteous, isaiah, merciful, reasonable
8. chrysanthemum, considerate, lynx, prevaricate, reuben
9. ezra, ichabod, ledger, parchesi, preceding
10. crocus, dahlia, jonquil, opossum, poltroon
- 10.5 begonia, equitable, pretentious, renegade, reprobate
11. armadillo, iguana, philanthropic

The child's score or measure is determined by finding the hardest, or the highest-numbered, line that he marks with not more than a single error — all omissions being regarded as errors. The number of this line is taken as the child's score. For example, five children gave results as follows:

NUMBER OF OMISSIONS AND ERRORS IN EACH LINE IN THE CASE  
OF FIVE PUPILS, C, J, N, R, AND W

Line . . . . .	4	5	6	7	8	9	10	10.5	11
Pupil C . . . . .	0	0	0	0	1	3	3	4	3
“ J . . . . .	0	0	0	0	1	1	1	3	2
“ N . . . . .	0	0	0	1	2	4	4	3	3
“ R . . . . .	0	0	0	0	1	2	3	5	3
“ W . . . . .	0	0	0	0	0	0	1	1	0

Thus we may say that C has ability 8; J has ability 10; N has ability 7; R has ability 8; W has ability 10.5 or 11 or better.

To measure the ability of a class as a whole, we simply take an average by adding together all the errors and omissions on each line and dividing by the number of children. In a rough estimate the class gets credit for the highest-numbered line that shows an average error of 1 or less, and the figure thus obtained, or the result as a whole, can be used for comparison with the achievement of other classes.

Considering the five children above mentioned as a class, we have, as the average number of errors and omissions on each line, the following:

Line. . . . .	4	5	6	7	8	9	10	10.5	11
Errors (including omissions) . .	0	0	0	.2	1.0	2.0	2.4	3.2	(2.2
									for the three-word line 1.1 or 3.7 for a five-word line of equal difficulty)

Since the highest-numbered line that this class marked with not more than an average of one error or omission per child is line number 8, 8 may be considered the score or measure of this class.

The choice of four out of five correct as a standard could be replaced by all correct (100%) or three correct (60%), but for statistical reasons 80% is the best criterion.

The measures procured by this method are not only objective, but they have a definite meaning. For instance, to say that an individual or class possesses ability 6 in reading, means that he or they possess the ability to mark correctly at least four out of five (80%) of the words in line 6 on the scale. Furthermore, the difference in difficulty between lines 4 and 5 is approximately equal to the difference in difficulty between lines 5 and 6, and so on. Lastly, since the difference in difficulty between lines 8 and 4 is probably about equal to that between 4 and 0, the attainment of a class scored 8 may be said to be about twice as great as that of one scored 4.

In the case of an individual, it is always possible to state in just what line are recorded errors and omissions totaling 20% or less (80% or more correct); that is, the individual's "degree of difficulty" can be accurately stated at sight. When it is a matter of an entire class, however, this is not so easy. For instance, a class may have a record of 16% of errors and omissions for line 6, and 25% for line 7. In such a case the "degree of difficulty" which would give a percentage of 20 may be obtained by consulting the tables and following the directions in the original paper. In short, when the percentage of errors or omissions for a given line, say 6, is known, it is possible to estimate just the "degree of difficulty," say 6.7, which would give a percentage of exactly 20 (80% correct) for the class in question.

The time required to measure a class of forty pupils, record the results, and estimate the "degree of difficulty"

that would give a percentage of 20 errors and omissions, varies from two to five hours. Thorndike believes it would be well for every school, from the fourth to the eighth grades, to make such measurements at the beginning and at the end of the school year.

This scale is not without defects and limitations, some of which Thorndike is at present working to overcome. (1) A scale whose steps or lines contain 10 or 20 words, instead of 5, will obviously give data for a more precise estimate of the ability of a class. (2) When applied to a single pupil the scale is not so precise as when applied to a class, for a child who happened to be interested in flowers and animals would have a decided advantage over one who was not so interested in them. (3) A pupil's score cannot always be exactly stated; for, if a child misses 2 words in line 8, no words in line 9 and 3 words in line 10, shall his ability be classed as 7 or 9? However, a reasonably rough estimate of his ability may be gained by consulting his score in the other lines. In a class, the chance familiarity of a pupil with certain words will be counterbalanced by the chance unfamiliarity of some other pupil. (4) The fact that words expressing relations, such as pronouns, prepositions, and conjunctions, are omitted in this scale, seems a serious limitation until it is considered that the chief importance of these words is in sentence comprehension, and that the scale for that purpose, which will be described later, tests knowledge of them rather thoroughly. (5) Not all the words on a given line are of absolutely equal difficulty, but the differences in the degree of difficulty are not of enough importance to constitute a defect. (6) It must be admitted that the differences between successive lines are not exactly equal. In fact, even "their approximate equality depends upon the approximate truth of certain hypotheses about the distribution of word-knowledge in children of the same grade and about the

comparative variability of the children in Grades IV, V, VI, VII, and VIII in respect to word-knowledge." (7) Lastly, it must be remembered that this scale does not measure the meaning of the printed words, save as required in the directions on the scale.

In spite of the difficulties which any such scale presents, it may be used for practical purposes, at its face value. It is capable of revealing large individual differences within a class and of measuring them roughly, if, as Thorndike says, it is interpreted with common sense. Moreover, as a measure of "the ability to understand printed words unconfused with the ability to express one's self orally or in writing," it is superior to any form of definition test. By extending it to include more difficult words it may be used to measure achievement and improvement from the third grade through college. Indeed, with slight modification it can be used to measure extent of vocabulary in any foreign language, and in fact, such scales for French, German, and Latin are being planned.

Scale A is designed for use in Grades IV to VIII inclusive in the elementary schools and to some extent in the high schools. To be sure that the general nature of the scale is understood, a short, simple, preliminary test, similar in character to Scale A, should be given. A pupil who has made less than five errors and omissions in the first two lines taken together in the preliminary test may be assumed to understand the general idea of the scale. A pupil in the third grade or above who makes more than ten errors and omissions in the first two lines taken together, may be assumed not to understand what is required of him. In the fourth grade one half an hour should be allowed for the test, in the fifth and sixth grades twenty-five minutes, and in the seventh and eighth grades, twenty minutes. Although a time record is not used in the measurement of the vocabulary itself,

Thorndike believes that it should be kept, without the pupil's knowledge, since it will prove instructive and requires little labor. A little experience will soon teach the scorer what lines he need score for a given class. For instance, in the eighth grade lines 4, 5, and 6 may almost always be neglected, while in the fourth grade lines 10, 10.5 and 11 may safely be disregarded.

The words of Scale A were chosen from a much larger number which were tried upon about 2500 pupils in the fourth, fifth, sixth, seventh and eighth grades in five different schools. Words were considered to be of approximately equal difficulty if approximately equal percentages of pupils in the fourth, fifth, sixth, seventh and eighth grades, respectively, marked them correctly in these tests. For instance, the words finally selected for row 4 in the scale — *camel*, *samuel*, *kind*, *lily*, and *cruel* — were marked correctly by approximately the same percentage of pupils in all the fifth grade classes, and similarly in the other grades; that is, about 100% got each of them right in the eighth grade, about 99% in the seventh grade, 98% in the sixth grade and 96% in the fifth grade. At present it is planned to improve the scale so that each row will include ten words instead of five. Some of the words added will be similar to those already in the scale, such as boys' names, while others will be words of equal difficulty obtained by administering new tests. An attempt will also be made to find words of difficulty 11.5, 12 and 12.5. With the material which he has already collected, Thorndike expects to enlarge Scale A by adding words of difficulty 4.5, 5.5, 6.5, 7.5, 8.5 and 9.5. In this way the exactitude of measurement of extent or range of visual vocabulary will be greatly increased.

---

(2) Scale Alpha. For Measuring the Understanding of Sentences

Thorndike's second scale, Scale Alpha, is an attempt to measure the ability of a child to read understandingly, that is, to understand the meaning of sentences and paragraphs. The value of such a scale is obvious when it is realized that competent judges would rate this ability "at from 60% to 90% of the total result to be sought by the elementary school in the teaching of reading."

In constructing this scale, preliminary experimentation was conducted along two separate lines; namely, (1) measurement by the passage-question method and (2) measurement by responses in marking letters, numbers, and the like. The work in both of these lines was so successful in measuring the pupil's ability to read understandingly, that the two types of measurement were employed in the final scale, which consists of four "sets" or steps, each one of which contains from one to five questions. This scale is reproduced below.

SET a OR 4

Read this and then write the answers. Read it again as often as you need to.

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All of these children except Will had red hair. He had brown hair.

1. Was John's sister tall or short?.....
2. How many brothers had John?.....
3. What was his sister's name?.....

## SET b OR 6

Read this and then write the answers. Read it again as often as you need to.

Long after the sun had set, Tom was still waiting for Jim and Dick to come. "If they do not come before nine o'clock," he said to himself, "I will go on to Boston alone." At half past eight they came bringing two other boys with them. Tom was very glad to see them and gave each of them one of the apples he had kept. They ate these and he ate one too. Then all went on down the road.

1. When did Jim and Dick come?.....
2. What did they do after eating the apples?....  
.....
3. Who else came besides Jim and Dick?.....  
.....
4. How long did Tom say he would wait for them?  
.....
5. What happened after the boys ate the apples?  
.....



## SET c OR 8

Read this and then write the answers. Read it again as often as you need to.

It may seem at first thought that every boy and girl who goes to school ought to do all the work that the teacher wishes done. But sometimes other duties prevent even the best boy or girl from doing so. If a boy's or girl's father died and he had to work afternoons and evenings to earn money to help his mother, such might be the case. A good girl might let her lessons go undone in order to help her mother by taking care of the baby.

1. What are some conditions that might make even the best boy leave school work unfinished?  
.....
2. What might a boy do in the evenings to help his family?.....
3. How could a girl be of use to her mother?....  
.....
4. Look at these words: *idle, tribe, inch, it, ice, ivy, tide, true, tip, top, tit, tat, toe.*

Cross out every one of them that has an *i* and has not any *t* (T) in it.

## SET d OR 10

Read this and then write the answers. Read it again as often as you need to.

It may seem at first thought that every boy and girl who goes to school ought to do all the work that the teacher wishes done. But sometimes other duties prevent even the best boy or girl from doing so. If a boy's or girl's father died and he had to work afternoons and evenings to earn money to help his mother, such might be the case. A good girl might let her lessons go undone in order to help her mother by taking care of the baby.

1. What is it that might seem at first thought to be true, but really is false? . . . . .
2. What might be the effect of his father's death upon the way a boy spent his time? . . . . .
3. Who is mentioned in the paragraph as the person who desires to have all lessons completely done? . . . . .
4. In these two lines draw a line under every 5 that comes just after a 2, unless the 2 comes just after a 9. If that is the case, draw a line under the next figure after the 5:

5 3 6 2 5 4 1 7 4 2 5 7 6 5 4 9 2 5 3 8 6 1 2 5

4 7 3 5 2 3 9 2 5 8 4 7 9 2 5 6 1 2 5 7 4 8 5 6

The foregoing scales and tables in this section are reproduced by the courtesy of Dr. E. L. THORNDIKE.

In the first two sets or steps — “a or 4” and “b or 6” — of the scale just given, the first type of measurement or the passage-question method, is used. Here the ability to understand a sentence or paragraph is measured by the correctness of verbal responses to certain questions asked regarding it. In the last two sets or steps — “c or 8” and “d or 10” — ability to understand a sentence or short paragraph is measured by responses which are not entirely verbal in character, such as marking letters and numbers. Each one of the four “sets” or steps is more difficult than the preceding one.

As in the case of Scale A, a preliminary test should be given the pupils before administering Scale Alpha, to find out if they understand instructions. Scale Alpha is available for Grades III to VIII. Twenty to thirty minutes should be allowed for administering it, and scoring is done as in Scale A. In marking the responses “the general intent should be to require an answer that proves that the pupil has understood the passage perfectly.” Because of the small number of steps in the scale, the “degree of difficulty” or, what amounts to the same thing, the ability of an individual or class, may be estimated from the percentage of errors and omissions nearest 20%, in a very similar manner to that used in Scale A; but for detailed directions the original paper must be consulted.

Thorndike points out that the values for the steps of this scale are not at all exact; that is, the difficulty of Set 4, for instance, is not exactly two and one-half times that of a possible Set 1, but he has permitted this estimate to stand, to facilitate the understanding of the scale. In order to obtain a scale of four or more steps, and make sure that all the questions in each step are of approximately equal difficulty and that there is a uniform difference in difficulty between the different steps, it will be

necessary to test over 4000 pupils, obtaining from each from 50 to 60 responses.

As Thorndike points out, even though Scale Alpha is but provisional, its use will make comparison much fairer and more exact than hours of oral questioning on the part of the most capable supervisor of reading. The scale will eventually be extended and improved by adding other elements equal in difficulty to those now given and by filling in with intermediate steps.

### **(3) The Gray Tentative Scale for Measuring Achievement in Oral Reading**

This provisional scale for measuring ability to pronounce English sentences consists of ten paragraphs of increasing reading difficulty.

#### **Passage a**

It was time for winter to come. The little birds had all gone far away. They were afraid of the cold. There was no green grass in the fields, and there were no pretty flowers in the gardens. Many of the trees had dropped all their leaves. Cold winter with its snow and ice was coming soon.

#### **Passage b**

Once there lived a king and queen in a large palace, but the king and queen were not happy. There were no little children in the house or garden. One day they found a poor little boy and girl at their door. They took them into the palace and made them their own. The king and queen were then happy.

## Passage c

Once I went home from the city for a summer's rest. I took my gun for a stroll in the woods where I had shot many squirrels. I put my gun against a tree and lay down upon the leaves. Soon I was fast asleep, dreaming of a group of merry, laughing children running and playing about me on all sides.

## Passage d

One of the most interesting birds which ever lived in my bird-room was a blue jay named Jakey. He was full of business from morning till night, scarcely ever still. He had been stolen from a nest long before he could fly, and he was reared in a house, long before he had been given to me as a pet.

## Passage e

The part of farming enjoyed most by a boy is the making of maple-sugar. It is better than blackberrying and almost as good as fishing. One reason he likes this work is that some one else does most of it. It is a sort of work in which he can appear to be very industrious, and yet do but little.

## Passage f

It was one of those wonderful evenings such as are found only in this magnificent region. The sun had sunk behind the mountains, but it was still light. The twilight glow embraced a third of the sky, and against its brilliancy stood the dull white masses of the mountains in evident contrast.

## Passage g

George Washington was in every sense of the word a wise, good and great man. But his temper was naturally irritable and high-toned. Through reflection and resolution he had obtained a firm and habitual ascendancy over it. If, however, it broke loose its bonds, he was most tremendous in his wrath.

## Passage h

Responding to the impulse of habit, Josephus spoke and the others listened attentively, but in grim and contemptuous silence. He spoke for a long time, continuously, persistently and ingratiatingly. Finally exhausted through lack of nourishment, he hesitated. As always happens in that contingency, he was lost.

## Passage i

The hypothesis concerning physical phenomena formulated by the early philosophers proved to be inconsistent and, in general, not universally applicable. Before relatively accurate principles could be established, physicists, mathematicians, and statisticians had to combine forces and work arduously.

## Passage j

Read the following sentences correctly: Sophistry is fallacious reasoning. They resuscitated him. Verbiage is wordiness. Equanimity is evenness of mind. He has a pertinacious, obstinate disposition. There was subtlety and poignancy in his remarks. A hypocritical or pharisaical nature is usually cynical.

The scale and table in this section are reproduced by the courtesy of Mr. W. S. GRAY.

## DIRECTIONS FOR ADMINISTERING

Pupils are required to read these passages in order, stopping at the end of each paragraph. The gross errors, minor errors, omissions, substitutions, and insertions made in each passage, as well as the time needed to read it, are recorded in detail on a duplicate of the scale. When a child makes 4 or more errors and takes 30 seconds or more to read a given paragraph, or when he makes 5 or more errors, however quickly he reads, he may be considered to have failed to read that passage. Although the difference in the degree of difficulty between any two of these passages has not as yet been definitely established, if values must be assigned to the ten paragraphs, Gray suggests that the following figures be used.

PASSAGE	VALUE	PASSAGE	VALUE
a	4.5	f	9.5
b	5	g	11
c	6	h	12
d	7	i	14
e	8	j	15

When finished, this scale will consist of an exactly graded series compiled from many graded series similar to the one just given. Even in this rough approximation to its final form, the scale is much better than any other means at hand for measuring ability in pronouncing English sentences.

## II. STARCH READING TESTS

## No. 1

Once there was a little girl who lived  
with her mother.

They were very poor.

Sometimes they had no supper.

Then they went to bed hungry.

One day the little girl went into the  
woods.

She wanted sticks for the fire.

She was so hungry and sad!

“Oh, I wish I had some sweet por-  
ridge!” she said.

“I wish I had a pot full for mother  
and me.

We could eat it all up.”

Just then she saw an old woman with  
a little black pot.

She said, “Little girl, why are you so  
sad?”

“I am hungry,” said the little girl.



## No. 2

Betty lived in the South, long, long ago. She was only ten years old, but she liked to help her mother.

She had learned to do many things. She could knit and sew and spin; but best of all she liked to cook.

One day Betty was alone at home because her father and mother and brother had gone to town to see a wonderful sight.

The great George Washington was visiting the South. He was going from town to town, riding in a great white coach trimmed with shining gold. It had leather curtains, and soft cushions. Four milk-white horses drew it along the road.

Four horsemen rode ahead of the coach to clear the way and four others rode behind it. They were all dressed in white and gold.

## No. 3

Little Abe hurried home as fast as his feet could carry him. Perhaps if he had worn stockings and shoes like yours, he could have run faster. But, instead, he wore deerskin leggings and clumsy moccasins of bearskin that his mother had made for him.

Such a funny little figure as he was, hurrying along across the rough fields! His suit was made of warm homespun cloth. His cap was made of coonskin, and the tail of the coon hung behind him, like a furry tassel.

But if you could have looked into the honest, twinkling blue eyes of this little lad of long ago, you would have liked him at once.

In one hand little Abe held something very precious. It was only a book, but little Abe thought more of that book than he would have thought of gold or precious stones.

You cannot know just what that book meant to little Abe, unless you are very fond of reading. Think how it would be to see no books except two or three old ones that you had read over and over until you knew them by heart!

## No. 4

The red squirrel usually waked me in the dawn, coursing over the roof and up and down the sides of the house, as if sent out of the woods for this very purpose.

In the course of the winter I threw out half a bushel of ears of sweet-corn on to the snow crust by my door, and was amused by watching the motions of the various animals which were baited by it. All day long the red squirrels came and went, and afforded me much entertainment by their maneuvers.

One would approach, at first, warily through the shrub-oaks, running over the snow crusts by fits and starts like a leaf blown by the wind. Now he would go a few paces this way, with wonderful speed, making haste with his "trotters" as if it were a wager; and now as many paces that way, but never getting on more than half a rod at a time.

Then suddenly he would pause with a ludicrous expression and a somerset, as if all eyes in the universe were fixed on him. Then, before you could say Jack Robinson, he would be in the top of a young pitch-pine, winding up his clock and talking to all the universe at the same time.

## No. 5

Once upon a time, there lived a very rich man, and a king besides, whose name was Midas; and he had a little daughter, whom nobody but myself ever heard of, and whose name I either never knew, or have entirely forgotten. So, because I love odd names for little girls, I choose to call her Marygold.

This King Midas was fonder of gold than anything else in the world. He valued his royal crown chiefly because it was composed of that precious metal. If he loved anything better, or half so well, it was the one little maiden who played so merrily around her father's footstool. But the more Midas loved his daughter, the more did he desire and seek for wealth. He thought, foolish man! that the best thing he could possibly do for his dear child would be to give her the immensest pile of yellow, glistening coin, that had ever been heaped together since the world was made. Thus, he gave all his thoughts and all his time to this one purpose. If ever he happened to gaze for an instant at the gold-tinted clouds of sunset, he wished that they were real gold, and that they could be squeezed safely into his strong box. When little Marygold ran to meet him, with a bunch of buttercups and dandelions, he used to say, "Poh, poh, child! If these flowers were as golden as they look, they would be worth the plucking!"

And yet, in his earlier days, before he was so entirely possessed of this insane desire for riches, King Midas had shown a great taste for flowers.

## No. 6

In a secluded and mountainous part of Stiria there was in old times a valley of the most surprising and luxuriant fertility. It was surrounded on all sides by steep and rocky mountains, rising into peaks which were always covered with snow, and from which a number of torrents descended in constant cataracts. One of these fell westward over the face of a crag so high that, when the sun had set to everything else, and all below was darkness, his beams still shone full upon this waterfall, so that it looked like a shower of gold. It was, therefore, called by the people of the neighborhood, the Golden River. It was strange that none of these streams fell into the valley itself. They all descended on the other side of the mountains, and wound away through broad plains and past populous cities. But the clouds were drawn so constantly to the snowy hills, and rested so softly in the circular hollow, that in time of drought and heat, when all the country round was burnt up, there was still rain in the little valley; and its crops were so heavy and its hay so high, and its apples so red, and its grapes so blue, and its wine so rich, and its honey so sweet, that it was a marvel to every one who beheld it, and was commonly called the Treasure Valley.

The whole of this little valley belonged to three brothers called Schwartz, Hans and Gluck. Schwartz and Hans, the two elder brothers, were very ugly men, with overhanging eyebrows and small, dull eyes.

## No. 7

Captain John Hull was the mint-master of Massachusetts, and coined all the money that was made there. This was a new line of business, for in the earlier days of the colony the current coinage consisted of gold and silver money of England, Portugal, and Spain. These coins being scarce, the people were often forced to barter their commodities instead of selling them.

For instance, if a man wanted to buy a coat, he perhaps exchanged a bearskin for it. If he wished for a barrel of molasses, he might purchase it with a pile of pine boards. Musket-bullets were used instead of farthings. The Indians had a sort of money called wampum, which was made of clamshells, and this strange sort of specie was likewise taken in payment of debts by the English settlers. Bank-bills had never been heard of. There was not money enough of any kind, in many parts of the country, to pay the salaries of the ministers, so that they sometimes had to take quintals of fish, bushels of corn, or cords of wood instead of silver or gold.

As the people grew more numerous and their trade one with another increased, the want of current money was still more sensibly felt. To supply the demand the general court passed a law for establishing a coinage of shillings, sixpences, and threepences. Captain John Hull was appointed to manufacture this money, and was to have about one shilling out of every twenty to pay him for the trouble of making them.

## No. 8

The years went on, and Ernest ceased to be a boy. He had grown to be a young man now. He attracted little notice from the other inhabitants of the valley; for they saw nothing remarkable in his way of life, save that, when the labor of the day was over, he still loved to go apart and gaze and meditate upon the Great Stone Face. According to their idea of the matter, it was a folly, indeed, but pardonable, inasmuch as Ernest was industrious, kind, and neighborly, and neglected no duty for the sake of indulging this idle habit. They knew not that the Great Stone Face had become a teacher to him, and that the sentiment which was expressed in it would enlarge the young man's heart, and fill it with wider and deeper sympathies than other hearts. They knew not that thence would come a better wisdom than could be learned from books, and a better life than could be molded on the defaced example of other human lives. Neither did Ernest know that the thoughts and affections which came to him so naturally, in the fields and at the fireside, and wherever he communed with himself, were of a higher tone than those which all men shared with him.

By this time poor Mr. Gathergold was dead and buried; and the oddest part of the matter was, that his wealth, which was the body and spirit of his existence, had disappeared before his death, leaving nothing of him but a living skeleton, covered over with a wrinkled, yellow skin. Since the melting away of his gold, it had been very generally conceded that there was no such striking resemblance, after all, betwixt the ignoble features of the ruined merchant and that majestic face upon the mountainside.

## No. 9

To an American visiting Europe, the long voyage he has to make is an excellent preparative. The temporary absence of worldly scenes and employments produces a state of mind peculiarly fitted to receive new and vivid impressions. The vast space of waters that separates the hemispheres is like a blank page in existence. There is no gradual transition, by which, as in Europe, the features and population of one country blend almost imperceptibly with those of another. From the moment you lose sight of the land you have left, all is vacancy until you step on the opposite shore, and are launched at once into the bustle and novelties of another world.

In traveling by land there is a continuity of scene and a connected succession of persons and incidents, that carry on the story of life, and lessen the effect of absence and separation. We drag, it is true, "a lengthening chain," at each remove of our pilgrimage; but the chain is unbroken: we can trace it back link by link; and we feel that the last still grapples us to home. But a wide sea voyage severs us at once. It makes us conscious of being cast loose from the secure anchorage of settled life, and sent adrift upon a doubtful world. It interposes a gulf, not merely imaginary, but real, between us and our homes — a gulf subject to tempest, and fear, and uncertainty, rendering distance palpable, and return precarious.

The tests and standard scores in this section are reproduced by the courtesy of Dr. Daniel Starch.



## DIRECTIONS FOR ADMINISTERING TESTS

The series of tests published by Starch are designed to measure (1) comprehension of material read, (2) speed of reading, and (3) correctness of pronunciation. These tests, nine in number, are actually a graded series of passages chosen from various graded readers, each of them bearing a number which indicates the grade from which it was taken and in which it is to be used. For example, No. 1 is to be used in the first grade; No. 2, in the second grade; and so on. It should be noted that full directions for administering the tests accompany them.

Explain to the pupils that they are to read silently as rapidly as they can and at the same time to grasp as much as they can, and that they will be asked to write down, not necessarily in the same words, as much as they will remember of what they read.

They should also be told not to read anything over, but to read on continuously as rapidly as is consistent with grasping what they read.

Use for a given grade the test blank that bears the same number as that grade. For example, use No. 4 with the fourth grade, No. 5 with the fifth grade, etc. On the next day repeat the test in the same manner, but use the blank of the grade next below yours; that is, in the fourth grade use No. 3, in the fifth grade use No. 4, etc.

The blanks for the test should be distributed to the pupils *with the backs of the blanks up*, so that no one will be able to read any of the material until all are ready. Then give the signal "turn" and "start." Allow them to read *exactly thirty seconds*. Then have the pupils make a mark with pencil after the last word read to indicate how far they had read.

Then have them turn the blanks over immediately and write on the back all that they remember having read.

Allow as much time as they need, but make sure that they do not copy from one another, or turn the blank over to see the text. Finally, have them fill out the spaces at the bottom of the blank.

Make sure of allowing exactly 30 seconds for the reading. See that all pupils start and stop at the same time.

Since selection No. 1 was taken from a typical first grade reader, selection No. 2 from a typical second grade reader, and so on, it was assumed that the increase in difficulty from one passage to another was fairly uniform. Nevertheless, Starch carefully examined all the data obtained from administering the tests to about 1400 pupils. These data indicated that the assumption was correct that the passages increase in difficulty with approximate uniformity from step to step. They also seemed to show that, unless the selections have been read shortly before by the pupils tested, the value of the tests is not affected by the fact that some of the selections are more or less familiar fables or pieces of literature.

### (1) Reading Comprehension Test

In using the test to measure reading comprehension, the pupil is given a limited time — thirty seconds — to read as much as he can of the selection. He is then required to write out as much as he can of what he has read. The exact amount of understanding shown is determined by counting the number of words written which correctly express the thought of the selection. All words which reproduce the ideas of the test passage incorrectly, all words expressing added ideas or repeated ideas, are crossed out, and the number of remaining words is reckoned as the measure of comprehension. For instance, if a pupil in reproducing test No. 8, a selection of 142 words, uses 77 words and 5 are crossed out, his score is 72.

Starch answers the objection to written reproduction as an index of comprehension, by saying that if it is a

handicap it is the same for all, since the pupil who is at a distinct disadvantage in writing, as compared with speaking, is either very rare or fictitious. Immediate reproduction was thought best because it does away with the memory factor and imposes uniformity. The immediate memory span of an adult in verbatim reproduction of words in sentences is 25 words, and that of a child of six about 12 words, but in the time allotted for the test, the average eighth grade child will read 120 words and the average first grade pupil will read 45 words. Therefore, the chance of a child's memorizing a great part of the passage is eliminated by the length of the selection.

Another possible way of testing comprehension is to measure the ability of a child to answer certain questions concerning the test passage. This method was actually tried, but the results from its use were less accurate and more difficult to score than those from the method finally adopted.

The method of scoring comprehension by counting the number of words written which correctly reproduce the thought of the test passage, was adopted because it is "simple, rapid and objective." Two other methods, that of assigning percentage marks and that of finding the number of ideas correctly expressed, might have been used; but the former was disregarded because of its subjective character, and the latter, because it involved the difficulty of determining just what an idea is. For instance, is "hurried" a separate idea, or should "hurried along" be considered as one?

## (2) Test for Speed of Reading

The speed of reading is easily measured by determining how much of a given test passage the child is able to read in thirty seconds. By using a blank on which the number of words is indicated line by line to the end of the passage, the total number of words read in a given

time may be seen at a glance. This number, divided by thirty, is the child's score per second. Thirty seconds was chosen as the time limit, first, because "the necessary text for this interval could all be printed on a sheet of paper about the size of an ordinary page in a reader; and second, because a longer interval of time would increase very materially the labor of scoring the results." To ascertain whether thirty seconds is a long enough interval to test a pupil's reading capacity, preliminary tests were made which showed that both speed and comprehension remain nearly constant, irrespective of the length of the passage.

### (3) Test for Correctness of Pronunciation

Correctness of pronunciation is measured by noting the number of words pronounced incorrectly. The test is administered after the other two tests are completed, and when the pupil has, in consequence, acquired a certain familiarity with the passage. Of course, this test must be given individually and out of the hearing of the other pupils.

To test the validity of these measurements of reading capacity, a comparison was made in a school of 256 pupils between efficiency in reading as shown by the tests and as indicated by marks in reading assigned by teachers. The relation between the results of the tests and the reading as estimated by the teachers was close.

### STANDARD SCORES IN READING

On the basis of the results from the administration of the tests to over 3500 children in 15 schools in 7 cities and 3 states, the following tentative standard scores of efficiency have been made for each grade.

GRADES	SPEED OF READING (words per second)	COMPREHENSION (words written)
1	1.5	15
2	1.8	20
3	2.1	24
4	2.4	28
5	2.8	33
6	3.2	38
7	3.6	45
8	4.0	50

These tests show that great individual differences exist among pupils in the same grade. For example, in one of the fourth grades tested, one pupil showed a speed of reading of .8 words per second and another, of 4.7. Since the standard for speed in the first grade is 1.5 words per second and in the eighth grade, 4.0 words per second, the former pupil falls considerably below the standard of the first grade and the latter rises above the standard of the eighth grade. The same holds true of comprehension. In combined speed and comprehension the best pupil in one fourth grade made a score four and one half times as high as the poorest.

This wide difference in ability in a single grade means that a large amount of overlapping exists between different grades. In fact, on the basis of the studies made so far with these tests, it may be said that "one third of the pupils of any given grade could do the reading work of the next grade above as well as the average of that grade, one fifth could do the work of the second grade above as well as the average of that grade, and one eighth could do the work of the third grade above as well as the average of that grade."

## III. COURTIS READING TESTS

Courtis has constructed two different reading tests, one to measure rate and retention in normal reading (Test No. 4, Normal Reading, Series C) and the other to measure rate and retention in careful reading (Test No. 5, Careful Reading, Series C).

*Rate of normal reading* is determined by telling the child to read a selected passage for one minute at his natural rate of reading. (This test will be found reproduced on pages 96 and 97.)

At the end of this time the pupil is to draw a circle around the last word he has read. Since the words are numbered to the end of the passage his rate of reading may be quickly determined.

*Retention in normal reading* is measured by giving the pupil a sheet of paper on which is the story that he has just read, but having in it here and there groups of three words (in parentheses), two of which words were not used in the original story. (See pages 98 and 99.)

He is to cross out the words which he does not remember seeing before, and, if he is unable to recall whether he has seen them or not, he is to cross them all out. He is to continue this until he comes to the word at which he stopped in the original story. For remembering to stop at this place he is given credit for one point.

Scoring is done by means of an Answer Card (see page 104) which gives the correct words used in the original story. This is placed beside the pupil's paper and every word which has been correctly crossed off is counted as a point. By adding these points the pupil's score in retention is obtained.

*The rate of careful reading* is determined in a similar manner. Retention, however, is measured by the amount of the selected passage that the child is able to reproduce. (See tests on pages 100 and 101.)

---

Scoring is done by means of an Answer Card (see pages 102 and 103) which contains a list of the points or main ideas in the passage. For the reproduction of each of these ideas the pupil is given a credit of one. His final score is the sum of these credits.

The following tests (pages 96 to 104) are reproduced by the courtesy of Mr. S. A. Courtis.

## English. Test No. 4. Normal Reading

Bessie's (*Adventures, Experiences, Story.*) Part B.

SCORE	
Number of	Words.....

No. of Words	
18	Before the frightened little girl could decide what to do, the dog sat up
32	on his hind legs and began to beg. He gave another impatient little "Bow!",
47	but this time his bark did not seem so sharp and terrifying to Bessie, and
60	her fears began to disappear. "Why, he's really a cute little doggie," she
73	thought, and sitting up, she timidly offered him her bowl. The dog needed
86	no second invitation, but eagerly lapped the milk until every drop was gone.
99	Bessie set the empty bowl down upon the doorstep and patted the little
113	dog's head. He in turn was equally friendly, wagging his tail and trying to
128	lick her hand. In a few minutes more the two were the best of friends,
142	racing about the garden in a wild game of tag. Sometimes it was Bessie,
156	laughing and screaming, who was chasing the dog this way and that in and
170	out among the bushes of the garden; sometimes it was the dog, barking and
183	jumping, who was chasing Bessie. More than once dog and girl were so
198	much in each other's way that both fell down, rolling over and over on the
212	soft grass. Not for a long time had Bessie had such a pleasant playmate.



They were on the front lawn now, resting a minute after a particularly  
wild romp. Suddenly, through the pickets of the fence, the dog spied a cat  
crossing the street outside. Immediately he dashed after her, squeezing  
between the pickets, and running down the street at top speed. "Here,  
doggie, doggie," the little girl cried, and ran after her playfellow, only to be  
stopped by the fence. From between the pickets, she could see both the cat  
and the dog rapidly disappearing down the street. Hot tears of disap-  
pointment welled from her eyes. She hurried to the gate and shook it; it  
was securely latched. Could she open it? Many times before had she  
tried, but without success. Little girls grow, however, and standing on the  
lower part of the gate, and stretching to her utmost, she was just able to  
press her little fingers against the latch. Click, and the gate swung open.  
Bessie was free to hurry down the street. The cat and dog had disap-  
peared around the corner of the next block.

Name ..... Grade .....

SCORE
Time.....
No. read.....
No. right.....

### Test No. 4—Normal Reading—Continued

Bessie's (*Experiences, Adventures, Story.*) Part B.

Before the (*terrified, frightened, poor*) little girl could (*decide, think, know*) what to do, the dog sat up on his hind legs and began to (*bark, tease, beg.*) He gave another impatient little (*"Bow!"*, *bark, whine*) but this time his bark did not seem so (*loud, sharp, harsh*) and terrifying to Bessie, and her (*fears, terror, distress*) began to disappear. "Why, he's really a (*nice, cute, dear*) little doggie," she thought, and sitting up, she (*cautiously, timidly, boldly*) offered him her bowl. The dog (*wanted, asked, needed*) no second invitation, but (*eagerly, quickly, rapidly*) lapped the milk until every drop was (*swallowed, eaten, gone*).

Bessie set the empty bowl down upon the (*grass, doorstep, ground*) and patted the little dog's (*head, back, neck*). He in turn was (*very, also, equally*) friendly, wagging his tail and trying to (*paw, kiss, lick*) her hand. In a few minutes more the two were the best of (*comrades, friends, playmates*) racing about the (*grounds, lawn, garden*) in a wild game of tag. Sometimes it was Bessie, laughing and (*screaming, dancing, skipping*) who was chasing the dog this way and that in and out among the (*trees, bushes, benches*) of the garden; sometimes it was the dog, barking and (*squeaking,*

1 3 4 5 6 7 9 10 11 12 13 15 16 17 18 19 20

jumping, biting) who was chasing Bessie. More than once dog and girl  
were so much in each other's way that (*they, both, each*) fell down, rolling  
over and over on the (*soft, green, cool*) grass. Not for a long time had  
Bessie had such a pleasant (*game, frolic, playmate*).

They were on the (*front, side, back*) lawn now, resting a minute after  
a particularly wild (*race, game, romp*). Suddenly, through the pickets of  
the fence, the dog (*noticed, saw, spied*) a cat crossing the street outside.  
Immediately he (*ran, dashed, chased*) after her, squeezing between the  
pickets, and running down the street at (*top, high, great*) speed. "Here,  
doggie, doggie," the little girl cried, and ran after her (*friend, playfellow,*  
*companion*) only to be stopped by the fence. From between the (*pickets,*  
*posts, palings*) she could see both the cat and the dog (*gradually, rapidly,*  
*slowly*) disappearing down the street. Hot tears of (*anger, grief, disappoint-*  
*ment*) welled from her eyes. She (*hurried, ran, went*) to the gate and  
shook it; it was (*safely, firmly, securely*) latched. Could she open it? Many  
times (*before, yesterday, halfheartedly*) she had tried, but without success.  
Little girls grow, however, and (*resting, standing, stepping*) on the lower  
part of the gate, and (*reaching, stretching, pushing*) to her utmost, she was  
just able to (*press, push, touch*) her little fingers against the latch. Click,  
and the gate (*came, was, swung*) open. Bessie was (*free, able, quick*) to  
hurry down the street. The cat and dog had disappeared around the corner  
of the next (*street, block, square*).

## English. Test No. 5. Careful Reading

## Bessie's Adventures. Part C.

SCORE
Number of Words -----

	No. of Words
Once out of the yard, Bessie ran straight down the street as fast as her two little legs would carry her. From the corner around which the dog and cat had disappeared, she could see, half way up the block of the cross street, a dog barking furiously at the foot of a tree. So Bessie ran in that direction, getting quite out of sight of her own home. When she drew near, however, she stopped running and stood gazing at the dog in dismay. Her cute little playfellow was nowhere to be seen. This dog was larger and darker, and his voice was loud and savage. What big teeth he had! How viciously he growled and snapped at the cat up the tree!	19 33 48 63 77 90 103 117 128

Suddenly a sense of loneliness oppressed the little girl. The world seemed big and strange; she must get back to her home and her mother. She turned and started to run again, making for the corner. A great auto	139 153 167
--	-------------------

truck thundered by, blowing its horn loudly. A group of children jeered  
at her, and a small boy tried to block her way. Bessie reached the corner  
at last but, confused, bewildered, she turned in the wrong direction.

On and on she ran, finding nothing but strange houses and unfamiliar  
scenes. A suspicion that she was lost flashed across her mind and filled it  
with terror. The suspicion grew to a certainty. She *was* lost! She would  
never see her mother any more! In a dumb agony of terror the little girl  
sank helplessly upon the nearest horseblock, then gave way to a storm of  
sobs she could no longer restrain.

179

194

205

217

231

244

259

272

278

## Test No. 5. Careful Reading

Points

## Bessie's Adventures. Part C.

Once out of the yard — Bessie ran (1) — straight down the street — as fast  
 as — her two little legs — would carry her. From the corner (2) — around  
 which — the dog and cat — had disappeared, — she could see (3) — half way up  
 the block — of the cross street — a dog — barking furiously — at the foot — of a tree.

So — Bessie ran — in that direction (4) — getting quite out of sight — of her  
 own home. (5) When she drew near — however — she stopped running — and  
 stood — gazing at the dog — in dismay (6). — Her cute little playfellow — was  
 nowhere to be seen. This dog — was larger and darker — and his voice — was  
 loud and savage. What big teeth he had! How viciously — he growled — and  
 snapped — at the cat — up the tree! (7)

Suddenly — a sense of loneliness — oppressed — the little girl. The world —  
 seemed — big and strange (8). She must — get back — to her home — and her  
 mother (9). She turned — and started to run again — making for the corner (10).  
 A great auto truck — thundered by — blowing its horn — loudly (11). A group

1  
2  
3  
4  
5  
6  
7  
8  
10  
11

of children — jeered at her — and a small boy — tried to block her way (12).  
 Bessie reached the corner—at last—but confused, bewildered (13)—she turned—  
 in the wrong direction (14).  
 12  
 13  
 14

On and on — she ran (15) — finding nothing but strange houses — and  
 unfamiliar scenes (16). A suspicion — that she was lost — flashed across her  
 mind — and filled it — with terror. The suspicion — grew to a certainty. She  
 was lost! (17) She would never see — her mother — any more. In a dumb  
 agony — of terror (18) the little girl — sank helplessly upon the nearest horse—  
 block — then gave way — to a storm of sobs — she could no longer restrain (19).  
 15  
 16  
 17  
 18  
 19

## POINTS

- |                                |   |                           |
|--------------------------------|---|---------------------------|
| 1. Bessie ran after animals    | 9. Decided to go home                       | 15. Ran on and on         |
| 2. Looked from corner          | 10. Ran towards corner                      | 16. Neighborhood strange  |
| 3. Saw dog                     | 11. Auto truck went by                      | 17. Realized she was lost |
| 4. Ran in that direction       | 12. Children in her way                     | 18. Became terrified      |
| 5. Got out of sight of home    | 13. She became confused                     | 19. Was overcome by       |
| 6. Recognized her mistake      | 14. Turned the corner in wrong<br>direction | terror                    |
| 7. Differences observed in dog |   |                           |
| 8. Bessie felt lonely.         |   |                           |

## Test No. 4 — Normal Reading

## Part B

Experiences, Terrified,	1. Adventures, story.	They,	22. Both, each.
Bark, tease,	2. Frightened, poor.	Game, frolic,	23. Soft, green, cool.
Loud,	3. Decide, think, know.	Race, game,	24. Playmate.
Nice,	4. Beg.	Noticed, saw,	25. Front, side, back.
Cautiously,	5. Bow, bark, whine.	Ran,	26. Romp.
Wanted, asked,	6. Sharp, harsh.	Friend,	27. Spied.
Swallowed,	7. Fears, terror, distress.	Gradually,	28. Dashed, chased.
eaten,	8. Cute, dear.	Anger, grief,	29. Top, high, great.
Grass,	9. Timidly, boldly.	Safely, firmly,	30. Playfellow, companion.
Very, also,	10. Needed.	Resting,	31. Pickets, posts, palings.
Paw, kiss,	11. Eagerly, quickly,	Reaching,	32. Rapidly, slowly.
Comrades,	12. Gone.	Came, was,	33. Disappointment.
Grounds, lawn,	13. Doorstep, ground.	Street,	34. Hurried, ran, went.
Trees,	14. Head, back, neck.	41. Free, able, quick.	35. Securely.
Squealing,	15. Equally.	42. Block, square.	36. Before, yesterday, halfheartedly.
	16. Lick.	43. Drawing a line around the last word read.	37. Standing, stepping.
	17. Friends, playmates.		38. Stretching, pushing.
	18. Garden.		39. Press, punch, touch.
	19. Screaming, dancing, skipping.		40. Swung.
	20. Bushes, benches.		41. Free, able, quick.
	21. Jumping, biting.		42. Block, square.
			43. Drawing a line around the last word read.

Be sure to add 1 to the total number of points read to cover remembering to draw a line around the last word read.



To summarize then, a teacher of the fourth, fifth, sixth, seventh or eighth grade may test the ability of pupils, (1) in the *understanding* of single words by using Thorndike's Scale A, (2) in the *comprehension* of material read, by using Thorndike's Scale Alpha, the Starch series of tests, or the Courtis tests, and (3) in the rate of reading, by using either the Starch or the Courtis tests, preferably the Starch.

Thorndike's Scales may be obtained by sending to Teachers College, Columbia University, New York, and the Starch Reading Tests, by sending to the author at the University of Wisconsin. The sheets on which the scales and tests appear contain full directions for their use.

In using Scale A the teacher should allow thirty minutes for the test in the fourth grade, twenty-five minutes in the fifth and sixth grades, and twenty minutes in the seventh and eighth grades. In administering Scale Alpha the teacher should allow from twenty to thirty minutes. In Scale A the pupil's score is the highest numbered line that he marks correctly without more than a single error. Scale Alpha is scored in a similar manner; that is, the pupil's score is the highest numbered step or set in which he has answered at least three of the four questions correctly.

In using the Starch Tests the teacher should send for the test blank that bears the same number as her grade; for example, No. 4 for the fourth grade, No. 5 for the fifth grade, etc. The speed of reading is obtained by determining the number of words read in thirty seconds. The pupil's score is determined by counting the number of words in his written reproduction which correctly express the thought of the selection read. Added and repeated words, as well as those which represent the ideas of the selection incorrectly, are not counted.

Folders or manuals, covering every phase of the testing, together with answer cards, must be procured with

the test sheets if the Courtis Tests are to be used. These may be obtained by sending to the Department of Co-operative Research, 82 Eliot Street, Detroit, Michigan.

To measure oral reading Gray's Scale may be used. In reading the paragraphs in the scale, which gradually increase in difficulty, the gross errors, minor errors, omissions, substitutions, and insertions made in each paragraph are recorded. If a child makes 4 or more errors in a paragraph and takes 30 seconds or more to read it, or if he makes 5 or more errors, however quickly he reads, he may be considered to have failed in that paragraph. This scale may be obtained by sending to Teachers College, Columbia University, New York City.

#### EXERCISES

1. Describe in detail the methods you would employ for measuring the reading ability, oral and silent, of thirty children of Grade V, using (a) the Thorndike and Gray Scale and (b) the Starch Scale.

2. How would you compare your class with one of the same grade in another school, using the Starch Scale? What conditions would you have to meet to make the comparison of the results valid?

3. How do the results obtained from the Thorndike Scale compare with those which the Starch Scale give?

4. Does there seem to be any relation between speed of reading and comprehension of material read?

5. What distinctions between oral and silent reading have the tests revealed?

6. Have the tests revealed any marked difference in the reading ability of boys and girls? Of children of different nationalities? Of children who have used different reading textbooks?

7. In what way may a teacher modify Scale A so as to use it to test knowledge in various subjects in the curriculum from the elementary grades through college?

8. When should a teacher stop drill in oral reading and devote all the time to drill in comprehension?

9. Have the tests revealed wide variations in the reading ability of the pupils in your class or a condition of more or less uniformity?

10. What are the shortcomings of the scales described in this chapter? How could these be remedied?

## CHAPTER V

### SPELLING SCALES

#### I. BUCKINGHAM SCALE

#### II. STARCH SCALE

#### III. AYRES SCALE

#### I. BUCKINGHAM SPELLING SCALE

THIS investigation, following the lead taken by the experimental investigation of the quality of handwriting and of composition, had as its object the development of a scale for the measurement of spelling ability; a scale which would no longer depend upon chance selection of words and upon subjective judgments of teachers, but which would be of general application and purely objective. The results were first published in 1913.

It is an obvious fact that there is a great difference in words as regards ease of spelling. Thus, we can select words of the very simplest, such as *the, as, when*, up to words of extreme difficulty which can only be spelled after long acquaintance. Theoretically, therefore, it is possible to arrange a series of words along a scale in such a way that they become more and more difficult. Furthermore, it might be possible to arrange these words at equal intervals along the scale, these intervals being determined by the difficulty of each word. If in addition to this we fix a zero point (by taking the simplest words and agreeing that failure to spell these words indicates absence of spelling ability), a scale may be constructed which will measure the spelling ability of any individual, and will measure

the difficulty of any word which has to be spelled. Not only can we measure the spelling ability of individuals in this way, but also of classes, schools, and school systems.

Such measurements will be independent of individual opinion. Spelling ability will be determined, not by an arbitrary list of words, picked at random by individuals who have no knowledge of their relative degrees of difficulty, but by means of words on the scale, which have been standardized as regards their difficulty, by the simple device of finding out what percentage of eighth grade children spelled them correctly. The school has always attached great importance to spelling ability; whether or not this ability is overestimated, does not need discussion here. Suffice it to say, that if the school takes as its aim the teaching of spelling, it is essential that some method be devised to measure the extent to which the aim is accomplished.

Dr. Rice, as early as 1897, tested the pupils in all grades from the fourth to the eighth inclusive in twenty-one school systems, using a list of words, which has since become known as the Rice Sentence Test. This list is given on the following page.

## RICE SENTENCE LIST

- |                |                |                   |
|----------------|----------------|-------------------|
| 1. running     | 30. writing    | 59. sensible      |
| 2. slipped     | 31. language   | 60. business      |
| 3. listened    | 32. careful    | 61. answer        |
| 4. queer       | 33. enough     | 62. sweeping      |
| 5. speech      | 34. necessary  | 63. properly      |
| 6. believe     | 35. waiting    | 64. improvement   |
| 7. weather     | 36. disappoint | 65. fatiguing     |
| 8. changeable  | 37. often      | 66. anxious       |
| 9. whistling   | 38. covered    | 67. appreciate    |
| 10. frightened | 39. mixture    | 68. assure        |
| 11. always     | 40. getting    | 69. imagine       |
| 12. changing   | 41. better     | 70. peculiar      |
| 13. chain      | 42. feather    | 71. character     |
| 14. loose      | 43. light      | 72. guarantee     |
| 15. baking     | 44. deceive    | 73. approval      |
| 16. piece      | 45. driving    | 74. intelligent   |
| 17. receive    | 46. surface    | 75. experience    |
| 18. laughter   | 47. rough      | 76. delicious     |
| 19. distance   | 48. smooth     | 77. realize       |
| 20. choose     | 49. hopping    | 78. importance    |
| 21. strange    | 50. certainly  | 79. occasion      |
| 22. picture    | 51. grateful   | 80. exceptions    |
| 23. because    | 52. elegant    | 81. thoroughly    |
| 24. thought    | 53. present    | 82. conscientious |
| 25. purpose    | 54. patience   | 83. therefore     |
| 26. learn      | 55. succeed    | 84. ascending     |
| 27. lose       | 56. severe     | 85. praise        |
| 28. almanac    | 57. accident   | 86. wholesome     |
| 29. neighbor   | 58. sometimes  |                   |

The method of scoring was of the simple type which is usually found in schools, *i.e.* a mark was given for each word correctly spelled, or a unit subtracted for each word misspelled. That is, all words were taken as equal measures of spelling ability. It should be noted that the foregoing list contains among other words, *disappoint*, *necessary*, *changeable*, *better*, *because*, *picture*. An examination of these six words shows at once that they are by no means of equal difficulty. This was conclusively proved by Thorndike, who made an actual test of these words on a group of fifth grade children. Thus, in the group that he measured, while 37% failed to spell *necessary*, the failures to spell *better*, *because*, and *picture*, were 3%, 1%, 0%, respectively. This clearly shows that it is erroneous to measure the score of the individual by giving equal value to each of these words. The pupil who scores, let us say, 95%, has spelled not only all the easy words in the list, but also a considerable number of the hard ones, whereas the pupil who gets 50% has failed in the hard words, and has obtained his mark merely by spelling the easy words. That is, as the score increases, the units really get greater and greater, for to spell the five hardest words represents a very different task from spelling the five easiest words, and yet both have the same effect on the score. In other words, studies of this type must always lack precision because of the inequality of the units which are employed. They are useful for giving a rough estimate of the abilities of various groups, but when it comes to asking questions, such as: How does the spelling ability of one class differ from another? — the figures which represent the results give no quantitative information, and are actually misleading. As the science of school measurement advances, such a state of affairs can hardly be tolerated. Exact quantitative measurements of spelling ability are required; such quantitative results can never be obtained so long as the fundamental error

is made, that one word is equal to another word in difficulty, unless this is proved to be the case by actual measurements of large groups. To correct this error was the purpose of Buckingham's study of spelling ability.

The study was confined to grades from the third to the eighth, inclusive, of elementary schools located in or near the city of New York. The schools drew such different classes of children that any conclusions derived as a result of the study can be taken as representative. In all, about 9000 pupils were tested, a number from which general results might be expected; a greater number of pupils would not have increased the accuracy of the results sufficiently to compensate for the additional labor.

In the first test a list of 270 words was used. This will be called the "original list." This list was selected from a larger list of 5000 words taken from two or more of five special books used by the author in his own school. These 270 words had to satisfy two requirements: (1) All of them had to be words in the speaking vocabulary of a third grade child, and (2) a considerable portion of the words had to be of sufficient difficulty to test the spelling ability of an eighth grade child. These words were then placed in a continuous passage, and the whole dictated to Grades III to VIII in one school and to Grades IV to VII in another school. The dictation was very slow, so that the time factor did not enter. In marking the papers only the 270 words were regarded, those that served to link the whole into a continuous passage being neglected. All the papers were marked by the same person and two measurements were recorded: (1) the number of times each word was correctly spelled in each grade, and (2) the percentage of the entire number of words each pupil spelled correctly in each grade. We shall confine ourselves to the first consideration, *i.e.* to the number of times each word was correctly spelled.

TABLE I

Figures Indicate Per Cent Correct

Table reads: *across* was spelled correctly in the third grade of School II by 17% of the pupils; in the fourth grade of School I by 60% of the pupils, and of School II by 40% of the pupils, etc.

GRADE . . . .	3d	4th		5th		6th		7th		8th
SCHOOL . . . .	II	I	II	I	II	I	II	I	II	II
<i>across</i> . . . .	17	60	40	76	58	90	79	98	87	93
<i>addition</i> . . . .	2	38	26	60	28	76	45	94	76	83
<i>almost</i> . . . .	16	62	41	73	65	88	75	80	81	87
<i>alphabet</i> . . . .	25	13	1	63	12	40	46	82	43	68
<i>arithmetic</i> . . . .	27	89	53	100	72	96	92	100	97	98
<i>bridge</i> . . . .	29	59	42	87	52	98	85	100	94	97
<i>button</i> . . . .	14	50	35	70	49	77	63	84	62	83
<i>choose</i> . . . .	6	25	10	37	31	62	37	67	55	65
<i>day</i> . . . .	97	100	98	96	100	100	99	100	100	100
<i>guess</i> . . . .	6	29	17	67	30	77	50	82	66	85
<i>handful</i> . . . .	36	47	33	46	19	76	33	75	63	57
<i>pshaw</i> . . . .	1	4	6	29	6	46	5	31	31	18
<i>tomato</i> . . . .	34	83	49	67	43	74	48	79	32	38
<i>too</i> . . . .	0	10	3	17	4	26	7	63	22	27
<i>whose</i> . . . .	17	49	15	40	29	47	10	57	59	66

Table I represents the typical results obtained from the various grades in the particular schools. Thus for example, *across* was spelled correctly in the third grade of school II by 17% of the pupils, and in the seventh grade of school I by 98%. On the basis of these scores a group of 100 words, here called the "selected list," was chosen from the original list of 270 words.

The basis upon which the "selected list" was chosen is as follows: Referring to Table I, it will be seen that the word *across* was spelled by 17% of the third grade children, which means that it was not too hard to serve as a test of their ability. By the time the seventh and eighth grades



were reached, it still served as a test of ability, for it failed to be spelled in the seventh and eighth grades by 13% and 7%, respectively. For this reason the word *across* was selected. *Almost* and *button* were chosen for the same reason. On the other hand, *addition*, which was spelled by only 2% of the third grade children, was discarded as too difficult, for 2% could spell it rightly by mere chance, which means that the word really serves as no test for the particular grade.

### Continuous Passage — 100 Selected Words

*Whose answer is ninety? If the janitor sweeps, he will raise a dust. You ought not to steal even a penny. Wait until the hour for recess to touch the button. Smoke was coming out of their chimney. Every afternoon the butcher gave the hungry dog a piece of meat. One evening a carriage was stopping in front of my kitchen. I wear a number thirteen collar. Guess what made me sneeze. Send me a pair of leather shoes. I do not know, but I am almost sure they are mine. My uncle bought my cousin a pretty watch for forty dollars. The soldier dropped his sword. Jack had a whistle and also twelve nails. The ocean does not often freeze. You should speak to people whom you meet. It takes only a minute to pass through the gate and across the road. Did you ever hear a fairy laugh? The American Indian had a saucer without a cup. Neither a pear nor a peach was at the grocery store to-day. Cut up a whole onion with a handful of beans. My piano lesson was easy. The animal ran into the road and straight against a tree. Give me another sentence which has the word "title" in it. I believe true friends like to be together instead of apart.*

These 100 selected words (printed in italics) were again put into sentences as shown (page 113) and were dictated later to five schools. Great care was taken to insure uniformity in the administration of the tests. Later 18 additional words were added, making a total of 118 words dictated. The extent to which each of these 118 words was spelled correctly in each grade in each school was determined. Using the data so collected, it was possible to select words which show a regular increase in difficulty, as we pass down from grade to grade. From these words two lists were then selected, each containing 25 words; these are referred to as the "first preferred list" and "second preferred list," as tabulated below.

## PREFERRED LIST

FIRST		SECOND	
1. even	14. minute	26. already	39. too
2. lesson	15. cousin	27. beginning	40. towel
3. only	16. nails	28. chicken	41. Tuesday
4. smoke	17. janitor	29. choose	42. tying
5. front	18. saucer	30. circus	43. whole
6. sure	19. stopping	31. grease	44. against
7. pear	20. sword	32. pigeons	45. answer
8. bought	21. freeze	33. quarrel	46. butcher
9. another	22. touch	34. saucy	47. guess
10. forty	23. whistle	35. tailor	48. instead
11. pretty	24. carriage	36. telegram	49. raise
12. wear	25. nor	37. telephone	50. beautiful
13. button		38. tobacco	

Considering these 50 words alone, Table II shows the percentage of children from the third to the eighth grade, who were able to spell each of the 50 words. Thus, *even* was spelled correctly by 59% of children in the third grade, 93% in the sixth, and 97% in the eighth grade.

TABLE II  
(Showing Standard Scores in Spelling)

WORDS	3D YR.	4TH YR.	5TH YR.	6TH YR.	7TH YR.	8TH YR.
1. even . . .	59%	79%	89%	93%	93%	97%
2. lesson . . .	37	72	83	91	94	96
3. only . . .	65	75	89	95	97	99
4. smoke . . .	46	69	85	94	96	99
5. front . . .	51	72	80	90	94	97
6. sure . . .	47	55	69	78	89	94
7. pear . . .	31	42	58	72	81	94
8. bought . . .	40	65	79	91	94	97
9. another . . .	36	43	78	86	94	96
10. forty . . .	49	62	65	72	83	87
11. pretty . . .	45	67	76	90	90	94
12. wear . . .	35	49	61	74	84	93
13. button . . .	32	52	61	73	74	87
14. minute . . .	26	38	62	77	86	92
15. cousin . . .	19	47	69	89	89	95
16. nails . . .	43	58	71	87	92	96
17. janitor . . .	19	42	58	81	81	90
18. saucer . . .	11	29	42	58	79	81
19. stopping . . .	27	39	55	71	76	84
20. sword . . .	13	46	57	78	86	93
21. freeze . . .	29	46	68	83	86	94
22. touch . . .	45	52	60	81	84	93
23. whistle . . .	22	55	56	64	75	85
24. carriage . . .	13	40	50	67	81	85
25. nor . . .	63	61	65	68	77	94
26. already . . .	16	42	43	62	44	77
27. beginning . . .	9	25	37	46	66	75
28. chicken . . .	49	70	83	90	96	99
29. choose . . .	22	34	48	60	65	82
30. circus . . .	20	39	50	72	75	95
31. grease . . .	11	18	37	35	42	57
32. pigeons . . .	7	29	41	57	70	82
33. quarrel . . .	15	39	53	75	86	94
34. saucy . . .	14	35	40	52	71	78
35. tailor . . .	38	55	70	75	81	84
36. telegram . . .	15	31	39	63	73	84
37. telephone . . .	8	35	48	67	83	87
38. tobacco . . .	12	39	60	75	88	96
39. too . . .	14	28	27	24	30	43
40. towel . . .	24	44	64	73	78	94
41. Tuesday . . .	46	70	67	80	87	91
42. tying . . .	44	58	70	68	76	87
43. whole . . .	17	43	64	78	84	90
44. against . . .	19	30	54	75	84	94
45. answer . . .	27	47	67	86	90	97
46. butcher . . .	33	59	69	85	90	97
47. guess . . .	20	32	49	67	77	85
48. instead . . .	32	48	62	86	87	91
49. raise . . .	21	54	67	84	93	94
50. beautiful . . .	10	52	70	85	94	96

In this way, Buckingham has provided a basis of comparison, which may be used by any teacher, as a method of testing the relative ability of different classes.<sup>1</sup>

#### DIRECTIONS FOR ADMINISTERING

The following instructions, which are essentially the same as those followed by Buckingham, may be given as regards the conduct of the test:

(1) Give all the words in sentences during one session, *i.e.* either in morning or afternoon of same day, except in classes below the fifth grade, where the material should be given in two periods separated by half an hour at least.

(2) Each sentence should be dictated, either as a whole or in part, as many times as may seem necessary to secure its complete understanding. This experiment is purely a test in spelling; it is not expected that the pupils should be subjected to the added difficulty of recalling the words dictated.

(3) Offer no explanation of separate words or sentences. If the meaning is not clear, repeat the sentence as a whole or in part.

(4) Do not ask the children to underline words, or otherwise call attention to the significant words of the sentences.

(5) After the children have written the sentences, read them again, and allow the pupils to insert words or make other corrections before finally collecting the papers.

These papers may now be collected for the whole class, and the percentage of pupils getting *any particular word* correct determined and compared with the table which has already been given. Of course no particular significance is attached to any single word; there is no one word which will test the spelling ability of a group.

<sup>1</sup> The tables in this section are reproduced by the courtesy of Dr. B. R. Buckingham.

When, however, 50 words are taken, which have been previously standardized, the manner in which these are spelled by any group of pupils will serve to give a quantitative idea of their spelling ability. Thus, if it is found by a teacher who is dealing with Grade V, that her average percentage for 50 words falls notably below the average given in the table for Grade V, there is every reason to suppose that there is something abnormal about the standing of that class, due to causes which might profitably be investigated.

Suppose, for example, that we are dealing with a fifth grade which contains 50 children, and we find that the word *another* is spelled correctly by 31 of the children. Reducing this to the percentage basis, the score of the class for this word is 62%. On reference to Buckingham's Table, we see that the average score of this grade for the word *another* is 78%, which means that the particular grade in question, as far as this word is concerned, was not equal to the average. The same procedure may be repeated with any of the other words in the list, and the average of all the percentages obtained. This figure may then be compared with the averages of the percentage for Grade V given for the particular words employed. It is necessary to use from 10 to 20 words in testing a grade, in order to avoid the danger of picking out one or two words upon which special drill might have been given. When 10 or 20 words are chosen at random from the list, this difficulty is obviated.

It may appear that some justification is required for this laborious study. The ordinary individual would be apt to take the attitude that the teacher's judgment would be just about as sound as the estimates arrived at by the foregoing process. As a matter of fact, the 50 words were ranked by 300 judges, most of them teachers. Naturally there was a general agreement between the teachers' judgments, and the relative order of the words

found as the result of experimental study. But with certain words, there was very great disagreement. Thus, the word *nor* when ranked by the teachers was given fifth place as regards ease of spelling. The actual records show that the children found it the sixteenth word as regards ease of spelling. Again, the word *button* was ranked ninth by the teachers, and thirty-first by the records which came from the pupils. This shows the unsatisfactoriness of relying on teachers' judgments. As long as those who are teaching do not know the relative difficulty of the words taught, how can they be expected to give the correct weight either in time or emphasis in their teaching?

Buckingham, in the latter half of his study, proceeds to construct a scale for the measurement of spelling efficiency, a scale which contains at one end words which, if they cannot be spelled, would indicate zero ability, and at the other end words which are very difficult for the average child in the grades to spell. By simple statistical methods and suitable assumptions he determined the interval between the words on the scale, the length of the interval being measured by the increase in difficulty as shown by the percentage of times it was correctly spelled. It would be impossible in the limits of this book to explain the method of derivation of the scale. Its interest is largely theoretical, and in its present form it could not be used with profit by the average teacher. It should, however, be borne in mind that such a measuring rod has been constructed even in a difficult function such as spelling.

---

II. STARCH SPELLING SCALE

A second method of measuring spelling ability has been devised by Starch, who worked quite independently of Buckingham. While this method lacks the statistical precision of Buckingham's study, in that it assumes (as far as the score is concerned) each word to be of equal difficulty, it is very straightforward and has many points to recommend its use in the classroom. The first object of the experiment was to obtain six lists of equal difficulty, each containing 100 words, representative of the entire non-scientific English vocabulary. This was accomplished by taking at random the first defined word of more than two letters on every even-numbered page in Webster's New International Dictionary. This made a total of 1,186 words. Every technical, psychological and obsolete word was then discarded, leaving 600 words. These were then arranged alphabetically in the order of size beginning with three-letter words, four-letter words, etc. This list was then divided into six lists of 100 words each, by choosing for the first list, the first, seventh, thirteenth, etc., word of the original list of 600 words. The second list was obtained in a similar manner by taking the second, eighth, and fourteenth word, etc.; and so on till the sixth list, which was formed by taking the sixth and twelfth word, and so on. The lists which resulted from this process are as follows:

## LIST I

- |           |              |                     |
|-----------|--------------|---------------------|
| 1. add    | 35. prism    | 69. commence        |
| 2. but    | 36. rogue    | 70. estimate        |
| 3. get    | 37. shape    | 71. flourish        |
| 4. low    | 38. steal    | 72. luckless        |
| 5. rat    | 39. swain    | 73. national        |
| 6. sun    | 40. title    | 74. pinnacle        |
| 7. alum   | 41. wheat    | 75. reducent        |
| 8. blow   | 42. accrue   | 76. standing        |
| 9. cart   | 43. bottom   | 77. venturer        |
| 10. cone  | 44. chapel   | 78. ascension       |
| 11. easy  | 45. dragon   | 79. dishallow       |
| 12. fell  | 46. filter   | 80. imposture       |
| 13. foul  | 47. hearse   | 81. invective       |
| 14. gold  | 48. laden    | 82. rebellion       |
| 15. head  | 49. milden   | 83. scrimping       |
| 16. kiss  | 50. pilfer   | 84. unalloyed       |
| 17. long  | 51. rabbit   | 85. volunteer       |
| 18. mock  | 52. school   | 86. cardinally      |
| 19. neck  | 53. shroud   | 87. connective      |
| 20. rest  | 54. starch   | 88. effrontery      |
| 21. spur  | 55. vanity   | 89. indistinct      |
| 22. then  | 56. bizarre  | 90. nunciature      |
| 23. vile  | 57. compose  | 91. sphericity      |
| 24. afoot | 58. dismiss  | 92. attenuation     |
| 25. black | 59. faction  | 93. fulminating     |
| 26. brush | 60. hemlock  | 94. lamentation     |
| 27. close | 61. leopard  | 95. secretarial     |
| 28. dodge | 62. omnibus  | 96. apparitional    |
| 29. faint | 63. procure  | 97. intermissive    |
| 30. force | 64. rinsing  | 98. subjectively    |
| 31. grape | 65. splashy  | 99. inspirational   |
| 32. honor | 66. torpedo  | 100. ineffectuality |
| 33. mince | 67. worship  |                     |
| 34. paint | 68. bescreen |                     |



## LIST II

- |           |              |                     |
|-----------|--------------|---------------------|
| 1. air    | 35. quill    | 69. covenant        |
| 2. cat    | 36. rough    | 70. eugenics        |
| 3. hop    | 37. shout    | 71. friskful        |
| 4. man    | 38. stick    | 72. luminous        |
| 5. row    | 39. swear    | 73. opulence        |
| 6. tap    | 40. trump    | 74. planchet        |
| 7. awry   | 41. whirl    | 75. reformer        |
| 8. blue   | 42. action   | 76. thorough        |
| 9. cast   | 43. bridle   | 77. watering        |
| 10. corn  | 44. charge   | 78. belonging       |
| 11. envy  | 45. driver   | 79. displayed       |
| 12. feud  | 46. finger   | 80. indention       |
| 13. game  | 47. heaven   | 81. mercenary       |
| 14. grow  | 48. legend   | 82. redevelop       |
| 15. home  | 49. motley   | 83. senescent       |
| 16. knee  | 50. portal   | 84. uncharged       |
| 17. look  | 51. recipe   | 85. whichever       |
| 18. mold  | 52. scrape   | 86. centennial      |
| 19. part  | 53. simple   | 87. constitute      |
| 20. ruin  | 54. strain   | 88. exaltation      |
| 21. take  | 55. weaken   | 89. invocate        |
| 22. tree  | 56. breaker  | 90. personable      |
| 23. well  | 57. congeal  | 91. strawberry      |
| 24. allay | 58. disturb  | 92. concentrate     |
| 25. blaze | 59. foreign  | 93. imaginative     |
| 26. buggy | 60. hogger   | 94. mathematics     |
| 27. clown | 61. meaning  | 95. selfishness     |
| 28. doubt | 62. onerate  | 96. collectivity    |
| 29. false | 63. provoke  | 97. marriageable    |
| 30. forth | 64. salient  | 98. agriculturist   |
| 31. grass | 65. station  | 99. quarantinable   |
| 32. house | 66. trample  | 100. relinquishment |
| 33. money | 67. abstract |                     |
| 34. paper | 68. bulletin |                     |

## LIST III

- |           |              |                     |
|-----------|--------------|---------------------|
| 1. art    | 35. razor    | 69. dominate        |
| 2. dry    | 36. saint    | 70. exchange        |
| 3. ice    | 37. smell    | 71. governor        |
| 4. mix    | 38. stock    | 72. manifest        |
| 5. run    | 39. swoop    | 73. osculate        |
| 6. top    | 40. twine    | 74. pleasure        |
| 7. back   | 41. white    | 75. revising        |
| 8. bond   | 42. barrel   | 76. traverse        |
| 9. chip   | 43. buckle   | 77. westward        |
| 10. crib  | 44. cotton   | 78. capitally       |
| 11. ever  | 45. engine   | 79. extremism       |
| 12. fire  | 46. flimsy   | 80. indicated       |
| 13. guilt | 47. helmet   | 81. monoplane       |
| 14. hack  | 48. lesser   | 82. repertory       |
| 15. hunt  | 49. ocular   | 83. stimulate       |
| 16. lace  | 50. potato   | 84. unlocated       |
| 17. main  | 51. relate   | 85. accidental      |
| 18. more  | 52. season   | 86. citizenize      |
| 19. pelt  | 53. single   | 87. contribute      |
| 20. sand  | 54. supply   | 88. expertness      |
| 21. tang  | 55. weight   | 89. locomotive      |
| 22. turn  | 56. captain  | 90. prevailing      |
| 23. wine  | 57. contour  | 91. symmetrize      |
| 24. amuse | 58. earnest  | 92. consolatory     |
| 25. blind | 59. fowling  | 93. incremental     |
| 26. catch | 60. inflate  | 94. penetrative     |
| 27. count | 61. measure  | 95. superintend     |
| 28. dress | 62. palaver  | 96. conterminous    |
| 29. fancy | 63. raising  | 97. naturalistic    |
| 30. freak | 64. seizing  | 98. artificiality   |
| 31. gross | 65. sulphur  | 99. re-examination  |
| 32. inlet | 66. trestle  | 100. sentimentalism |
| 33. muddy | 67. adhesive |                     |
| 34. peace | 68. buttress |                     |

## LIST IV

- |           |               |                     |
|-----------|---------------|---------------------|
| 1. bee    | 35. remit     | 69. enabling        |
| 2. elk    | 36. scale     | 70. external        |
| 3. key    | 37. speak     | 71. greeting        |
| 4. new    | 38. stone     | 72. mosquito        |
| 5. saw    | 39. thick     | 73. outfling        |
| 6. war    | 40. under     | 74. positive        |
| 7. base   | 41. widen     | 75. romantic        |
| 8. book   | 42. bearer    | 76. undulate        |
| 9. clue   | 43. canine    | 77. adverbial       |
| 10. down  | 44. create    | 78. carpentry       |
| 11. fall  | 45. eraser    | 79. franchise       |
| 12. flat  | 46. garret    | 80. infatuate       |
| 13. girt  | 47. hollow    | 81. promenade       |
| 14. hand  | 48. little    | 82. rigmarole       |
| 15. iron  | 49. office    | 83. stripling       |
| 16. lime  | 50. prince    | 84. vegetable       |
| 17. make  | 51. retain    | 85. assignment      |
| 18. move  | 52. settle    | 86. comparison      |
| 19. plug  | 53. sluice    | 87. coördinate      |
| 20. shop  | 54. swerve    | 88. expressage      |
| 21. tear  | 55. withal    | 89. mayonnaise      |
| 22. tusk  | 56. chicken   | 90. recompense      |
| 23. wire  | 57. counter   | 91. untraveled      |
| 24. apple | 58. emperor   | 92. consumptive     |
| 25. blood | 59. freight   | 93. infuriation     |
| 26. chain | 60. journal . | 94. photosphere     |
| 27. craft | 61. neglect   | 95. terrestrial     |
| 28. drawn | 62. passion   | 96. horsemanship    |
| 29. field | 63. reserve   | 97. regenerative    |
| 30. frost | 64. serpent   | 98. circumscribed   |
| 31. guard | 65. surface   | 99. sculpturesque   |
| 32. jelly | 66. trouble   | 100. verisimilitude |
| 33. ocean | 67. affected  |                     |
| 34. pitch | 68. calendar  |                     |

## LIST V

- |           |              |                      |
|-----------|--------------|----------------------|
| 1. bow    | 35. revel    | 69. entirely         |
| 2. fly    | 36. scorn    | 70. farewell         |
| 3. law    | 37. spire    | 71. incident         |
| 4. old    | 38. strut    | 72. mountain         |
| 5. see    | 39. three    | 73. parallel         |
| 6. ache   | 40. voice    | 74. prelimit         |
| 7. bead   | 41. wince    | 75. spectral         |
| 8. call   | 42. beaver   | 76. urbanize         |
| 9. cold   | 43. cannon   | 77. aggrieved        |
| 10. draw  | 44. crispy   | 78. clarifier        |
| 11. fast  | 45. escape   | 79. hydraulic        |
| 12. foil  | 46. gladly   | 80. inheritor        |
| 13. glue  | 47. hustle   | 81. purgation        |
| 14. hard  | 48. mallet   | 82. sacrifice        |
| 15. jack  | 49. oriole   | 83. surviving        |
| 16. line  | 50. pulley   | 84. vestibule        |
| 17. mark  | 51. rubric   | 85. authorship       |
| 18. musk  | 52. shears   | 86. concoction       |
| 19. prig  | 53. solace   | 87. derigation       |
| 20. slat  | 54. trifle   | 88. federative       |
| 21. test  | 55. yellow   | 89. memorandum       |
| 22. vend  | 56. circuit  | 90. regularity       |
| 23. wood  | 57. crooked  | 91. abnormality      |
| 24. armor | 58. enstamp  | 92. disseminate      |
| 25. boast | 59. general  | 93. insensitive      |
| 26. chase | 60. lateral  | 94. predominate      |
| 27. cross | 61. nourish  | 95. unprevented      |
| 28. enjoy | 62. placard  | 96. inarticulate     |
| 29. fixed | 63. resolve  | 97. stupendously     |
| 30. glean | 64. signify  | 98. communicating    |
| 31. guild | 65. tabloid  | 99. anthropometric   |
| 32. joint | 66. unitive  | 100. emancipationist |
| 33. order | 67. approved |                      |
| 34. point | 68. cerebral |                      |

## LIST VI

- |           |              |                       |
|-----------|--------------|-----------------------|
| 1. box    | 35. river    | 69. erosible          |
| 2. gap    | 36. shaft    | 70. fetching          |
| 3. lay    | 37. stall    | 71. juncture          |
| 4. pod    | 38. sugar    | 72. narcotic          |
| 5. sex    | 39. throw    | 73. parasite          |
| 6. alms   | 40. watch    | 74. probator          |
| 7. bird   | 41. young    | 75. squeaker          |
| 8. camp   | 42. begird   | 76. vagabond          |
| 9. comb   | 43. causal   | 77. amphibian         |
| 10. dusk  | 44. discus   | 78. clearness         |
| 11. fear  | 45. ferret   | 79. impatient         |
| 12. foot  | 46. gutter   | 80. intestine         |
| 13. goat  | 47. killed   | 81. quadruple         |
| 14. hawk  | 48. middle   | 82. sauciness         |
| 15. keep  | 49. paddle   | 83. ticketing         |
| 16. life  | 50. puzzle   | 84. virulence         |
| 17. mass  | 51. sample   | 85. bafflement        |
| 18. navy  | 52. shield   | 86. condescend        |
| 19. raft  | 53. spring   | 87. disconcert        |
| 20. some  | 54. tubule   | 88. illiterate        |
| 21. that  | 55. bicycle  | 89. metropolis        |
| 22. vice  | 56. commode  | 90. repression        |
| 23. work  | 57. discard  | 92. animalcular       |
| 24. aside | 58. excuser  | 92. divestiture       |
| 25. brawn | 59. gravity  | 93. intrinsical       |
| 26. chime | 60. leaping  | 94. prerogative       |
| 27. crown | 61. obloquy  | 95. upholsterer       |
| 28. equip | 62. pontiff  | 96. interference      |
| 29. flock | 63. retreat  | 97. subantarctic      |
| 30. grand | 64. society  | 98. convocational     |
| 31. hedge | 65. tigress  | 99. imperturbation    |
| 32. knock | 66. vitiate  | 100. irresponsibility |
| 33. ought | 67. auditory |                       |
| 34. poppy | 68. churlish |                       |

These scales are reproduced by the courtesy of Dr. Daniel Starch.

The advantages of this method of selection are: (1) It gives a random sampling of the entire non-technical English vocabulary, for easy words and very hard words occur in the same proportion in the lists as in the English language. (2) The list contains words sufficiently easy to test the poorest speller. (3) The essential requirement of every scientific experiment is fulfilled, since another 600 words of the same average difficulty can be chosen, by employing the same method of selection, *e.g.* the tenth word in the dictionary could be used in place of the first word.

#### DIRECTIONS FOR ADMINISTERING TESTS

First have the pupils write the name, grade, school, city and date at the top of the sheet.

Pronounce the words clearly, but do not sound them phonetically, or inflect them so as to aid the pupils. Give the meaning of words that sound like words with a different meaning and spelling. The pupils are to write the words and to number them in the order in which they are given. Allow sufficient time for the writing.

Each grade is to be tested twice on two successive days. Use any one of the six lists on the first day and a different list on the second day. (When an entire school is being tested it may be desirable, though not necessary, to use on the first day the same list, say List 1, in all grades, and any other list on the second day.)

In the first grade use the first 40 words of the list, in the second grade use the first 65 words, in the third grade use the first 80 words, in the fourth grade use the first 90 words, and in all other grades use the entire list.

It has been demonstrated by administering the lists in schools, that each of them is of approximately the same difficulty. It is perhaps desirable, however, when measuring the efficiency of an individual group, to give two

of the tests. The average of the score made in the two tests will represent pretty accurately the spelling ability.

#### STANDARDS OF EFFICIENCY IN SPELLING

These spelling tests have been standardized by administering them to 2500 pupils in 12 schools of 5 cities, located in Wisconsin, Minnesota, and New York. The average results obtained are shown in the table below, in which the scores are given in round figures.

STANDARD SCORES FOR SPELLING

Grades . . .	1	2	3	4	5	6	7	8
	10	30	40	51	61	71	78	85

This table shows that on the average in Grade III in the schools measured, 40% of each list was spelled correctly. The point of most importance for the individual teacher is to know how the pupils of a particular grade compare in spelling efficiency with pupils of the same grade of other schools.

By using this very simple device, a purely objective measure of spelling ability can be obtained by the ordinary teacher. No longer need we speak of "good spellers," "bad spellers" and "medium spellers"; we can assign a numerical value to the spelling ability of each individual.

#### III. THE AYRES SPELLING TEST (1000 WORDS)<sup>1</sup>

Ayres has also presented a further method of measuring spelling ability based on the one thousand most common words in the English language. These words were chosen by combining the results of four previous investigations which had as their object the selection of the words most commonly used in different sorts of writing. The first study was founded on passages from the Bible and other well-known writings, including in all about 100,000

<sup>1</sup> The Ayres Spelling Scale (see insert) is reproduced by the courtesy of Dr. Leonard P. Ayres.

words. The second study of the frequency of different words was made on the basis of an analysis of the words used in 250 different articles taken from issues of four Sunday newspapers published in Buffalo. These articles, counting repetitions, contained 43,989 words; without repetitions, 6000 words. The third study consisted of the tabulation of 23,629 words from 2000 short letters written by 2000 people. The last study comprised a tabulation of some 200,000 words taken from the family correspondence of thirteen adults.

The list of 1000 words finally selected was determined by combining the results of all these studies. Thus, the 1000 words chosen were those which occurred most frequently in passages selected from a wide variety of sources; namely, the Bible, the writings of famous authors, newspaper articles, and private correspondence.

The method employed in standardizing the difficulty of each of the 1000 words was essentially the same as that used by Buckingham, but on a more extensive scale. The 1000 words were first made into 50 lists of 20 words each, and these lists were then administered, in the middle of the school year, to various grades in the schools of 84 cities scattered throughout the United States. The data secured from these tests made an aggregate of 1,400,000 spellings by 70,000 children. It was on the basis of this data that the Ayres Scale was constructed.

The scale presented explains itself. All the words in any particular column are of approximately the same spelling difficulty, the difficulty of each word having been determined by the percentage of times the word was spelled correctly in the tests mentioned above.

#### DIRECTIONS FOR ADMINISTERING

The details for administering the tests will be clear from the following example. Suppose we wished to measure



the spelling ability of any fifth grade. Taking any one of the columns given in the scale — say Column O — we would first of all select any twenty words from it. Then we would dictate these words in a list to the class, giving ample time for each word and explaining the meaning of a word, if doubtful, by putting it in a sentence. Lastly, we would collect the papers and calculate the number of words spelled correctly. If there were 30 children in the class, that would mean that 600 spellings were performed. Suppose out of these 600 spellings there were 480 correct. Then 80% of the words would be correctly spelled. A reference to the scale, Column O, shows that the fifth grade average at midyear is 84%, and the fourth grade average, 73%. Therefore the class measured would be a little below the average fifth grade standing. Suppose a particular child in the grade gets 18 correct out of the 20 words. This means a score of 90%, or slightly below the average for the sixth grade, which is 92%. The only care that must be taken in administering the test is not to select a list of words so short that there is a chance of not obtaining representative results. For this reason, in testing the ability of a particular pupil it is well not to use less than 20 words; but if a group is being tested, so as to obtain merely the group average, a smaller number of words may be used.

It should be noted that the standards published with the Ayres Scale only apply where these words have been given to pupils who have had no especial drill on them. For, since the words in the scale are so common that they form an excellent foundation for spelling, it is reasonable to suppose that special attention will be given them. This drill will make the pupil too familiar with them to have his score judged by the standard score as obtained by Ayres. This means that probably it will be necessary for each school to establish its own standards.

## EXERCISES

1. Select 15 words from the Buckingham Scale and use these for measuring the spelling ability of a particular class. Outline the steps you would take, and the way in which you would administer the test, score the papers, and tabulate the results.

2. What are the advantages derived from knowing the relative difficulties of different words? How should this alter the method of teaching?

3. Using the Starch Scale, how would you establish norms for the grades of your own school? Is it fair to expect a foreign district school and an English-speaking district school to produce the same percentages?

4. Suppose a teacher took any list of 100 words and administered these to a grade and discovered that on the average 75% of the spellings were correct, what would this tell or fail to tell the teacher?

5. If it was found that the average scores of a grade V, for successive years, tested in January on the Starch Scale, were 59, 60, 61, 62, 60, and the average fell suddenly to 53, where would you look for the cause?

6. How, by means of these scales, would it be possible to compare two different methods of teaching spelling?

7. If it is found that some children are very much better than the average for their grades, how should this affect the amount of time they devote to spelling? What should be done for those who are much poorer than the average?

8. Use (a) the Buckingham Scale, (b) the Ayres Scale, (c) the Starch Scale, to test the same class on successive days. Do the results agree, in that they show that the class has the same ability, measured by the grade norms?

9. Why would it not be fair to apply any of these tests if the children had been drilled on the lists used in these tests? Which is the safest scale to use if we wish to eliminate this error?

10. Administer List 1 and List 2 of the Starch Scale to the same class, on successive days, and compare the average scores in each. Should they be the same? Why?

## CHAPTER VI

### COMPOSITION SCALES

#### I. HILLEGAS SCALE

#### II. HARVARD-NEWTON SCALES

THE task of evaluating efficiency in composition is obviously a complex one because, not only are there several distinct types of composition, such as narration, description, etc., but merit in each of these types is the resultant of many independent factors. Attempts to estimate this efficiency — the qualities desirable in English composition — have resulted in the production of three separate methods of measuring.

The first method is that of the Hillegas Scale of mixed types of composition. This scale consists of a number of samples of English composition representing various types and ranging from very good to very poor in quality, each grade in the scale being represented by but one composition. For example, the sample composition representing one grade may be of the narration type, while that representing another grade may be of the description type. Since the composition to be measured is compared directly with the compositions in the scale, as in the Thorndike Handwriting Scale, the accurate comparison of one style of composition with an entirely different style, as is often necessary, is exceedingly difficult.

It was to do away with this objection that the second method of measurement, namely, the Harvard-Newton series of four scales, was formed. These scales measure efficiency in description, narration, exposition and argumentation, respectively.

Thirdly, there is the method originated by Rice and used with apparent success by Bliss and Courtis. Here no attempt is made to construct an actual scale; but progress in composition writing in an individual, class, or school is determined by simply noting the improvement shown by the individual, class, or school, in successive reproductions of similar selections at intervals throughout the school year. No attempt is made to express the value of the composition in per cents or otherwise. It is simply read, and placed in the class "Excellent," "Good," "Poor," etc., on the basis of the general impression produced by reading it. These initial attempts are so lacking in the precision for which the whole movement for standardization of school products stands, that they need no further description.

#### I. HILLEGAS COMPOSITION SCALE

The "Hillegas Scale for the Measurement of Quality in English Composition by Young People" consists of ten sample compositions which have been arranged in order of increasing merit, merit meaning that quality which competent persons consider as such. These samples have been assigned the following values: 0, 18, 26, 37, 47, 58, 67, 77, 83, and 93, respectively. These values are not based on the ordinary percentage system used in grading and should not be confused with such per cents. Instead, each one of the values represents the number of units of quality possessed by the composition to which it is attached. Thus, the composition rated 93 is approximately twice as good as the one rated 47, while the one rated 18 is approximately half as good as the one rated 37.

## 0

Dear Sir: I write to say that it aint a square deal Schools is I say they is I went to a school. red and gree green and brown aint it hito bit I say he don't know his business not today nor yesterday and you know it and I want Jennie to get me out.

## 18

the book I refer to reach is Ichabod Crane, it is an grate book and I like to rede it. Ichabod Crame was a man and a man wrote a book and it is called Ichabod Crane i like it because the man called it ichabod crane when I read it for it is such a great book.

## 26

Advantage evils are things of tyranny and there are many advantage evils. One thing is that when they opress the people they suffer awful I think it is a terriable thing when they say that you can be hanged down or trodden down without mercy and the tyranny does what they want there was tyrans in the revolutionary war and so the throwed off the yok.

## 37

## Sulla as a Tyrant

When Sulla came back from his conquest Marius had put himself consul so sulla with the army he had with him in his conquest seized the government for Marius and put himself in consul and had a list of his enemys printy and the men whoes names were on this list we beheaded.

## 47

## De Quincy

First: De Quincys mother was a beautiful woman and through her De Quincy inhereted much of his genius.

His running away from school enfluenced him much as he roamed through the woods, valleys and his mind became very meditative.

The greatest enfluence of De Quincy's life was the opium habit. If it was not for this habit it is doubtful whether we would now be reading his writings.

His companions during his college course and even before that time were great enfluences. The surroundings of De Quincy were enfluences. Not only De Quincy's habit of opium but other habits which were peculiar to his life.

His marriage to the woman which he did not especially care for.

The many well educated and noteworthy friends of De Quincy.

## 58

## Fluellen

The passages given show the following characteristic of Fluellen: his inclination to brag, his professed knowledge of History, his complaining character, his great patriotism, pride of his leader, admired honesty, revengeful, love of fun and punishment of those who deserve it.

## 67

## Ichabod Crane

Ichabod Crane was a schoolmaster in a place called Sleepy Hollow. He was tall and slim with broad shoulders, long arms that dangled far below his coat sleeves. His feet looked as if they might easily have been used for shovels. His nose was long and his entire frame was most loely hung to-gether.

## 77

## Going Down with Victory

As we road down Lombard Street, we saw flags waving from nearly every window. I surely felt proud that day to be the driver of the gaily decorated coach. Again and again we were cheered as we drove slowly to the postmasters, to await the coming of his majestie's mail. There wasn't one of the gaily bedecked coaches that could have compared with ours, in my estimation. So with waving flags and fluttering hearts we waited for the coming of the mail and the expected tidings of victory.

When at last it did arrive the postmaster began to quickly sort the bundles, we waited anxiously. Immediately upon receiving our bundles, I lashed the horses and they responded with a jump. Out into the country we drove at reckless speed — everywhere spreading like wildfire the news, "Victory!" The exilation that we all felt was shared with the horses. Up and down grade and over bridges, we drove at breakneck speed and spreading the news at every hamlet with that one cry "Victory!" When at last we were back home again, it was with the hope that we should have another ride some day with "Victory."

## 83

## Venus of Melos

In looking at this statute we think, not of wisdom, or power, or force, but just of beauty. She stands resting the weight of her body on one foot, and advancing the other (left) with knee bent. The posture causes the figure to sway slightly to one side, describing a fine curved line. The lower limbs are draped but the upper part of the body is uncovered. (The unfortunate loss of the statute's arms pre-

vents a positive knowledge of its original attitude). The eyes are partly closed, having something of a dreamy languor. The nose is perfectly cut, the mouth and chin are moulded in adorable curves. Yet to say that every feature is of faultless perfection is but cold praise. No analysis can convey the sense of her peerless beauty.

## 93

## A Foreigner's Tribute to Joan of Arc

Joan of Arc, worn out by the suffering that was thrust upon her, nethertheless appeared with a brave mien before the Bishop of Beauvais. She knew, had always known that she must die when her mission was fulfilled and death held no terrors for her. To all the bishop's questions she answered firmly and without hesitation. The bishop failed to confuse her for heresy, bidding her recant if she would live. She refused and was lead to prison, from there to death.

While the flames were writhing around her she bade the old bishop who stood by her to move away or he would be injured. Her last thought was of others and De Quincy says, that recant was no more in her mind than on her lips. She died as she lived, with a prayer on her lips, and listening to the voices that had whispered to her so often.

The heroism of Joan of Arc was wonderful. We do not know what form her great patriotism took or how far it really led her. She spoke of hearing voices and seeing visions. We only know that she resolved to save her country, knowing though she did so, it would cost her her life. Yet she never hesitated. She was uneducated save for the lessons taught her by nature. Yet she led armies and crowned the dauphin, king of France. She was only a girl, yet she could silence a great bishop by words that came from her heart and from her faith. She was only a woman, yet she could die as bravely as any martyr who had gone before.

**This scale is reproduced by the courtesy of Dr. M. B. Hillegas.**



The scale was derived in the following manner. The first step taken was the collection from various sources of about 7000 English compositions ranging from the very poorest to the best work done in the elementary and high schools. After these compositions had each been given a number from 1 to 7000, they were roughly graded by Hillegas and an assistant into ten classes, and from these ten classes 75 samples were selected. In order to have samples at both extremes of the scale, some artificial ones were supplied. Those placed at the zero end of the scale were conscious efforts by adults to write very poor English, while those placed at the one hundred end were obtained from youthful writings of certain literary geniuses and from the work of some college freshmen. As augmented, the set consisted of 83 samples varying from the poorest to the best by small degrees of quality. That the character of the handwriting might not influence the judges, all the samples were typewritten and mimeographed.

Separate sets of these samples were then sent to about 100 individuals, who were asked to arrange the samples in the order of their merit as specimens of English composition, calling the poorest specimen No. 1, the next, No. 2, and so on. Owing to the small number of judgments it was not possible to establish the position of any one sample with reasonable accuracy, but those samples that were of about equal merit were indicated. This resulted in the selection of a smaller set which still contained all the important steps in quality from the worst to the best.

This smaller set, comprising 27 samples, was selected by taking successively each of the samples in the larger group that about 75% of the judges had agreed was better than the last one selected. This percentage of judgments was taken for statistical reasons which will be explained later. Where large differences in merit existed

between two successive samples, new samples, judged by a number of individuals as ranging in merit between them, were introduced.

Then, as with the first set of samples, more than 100 of these sets consisting of 27 samples were mailed to competent critics of English literature, such as teachers, authors, and literary workers, with the request to rank them in order of literary merit. When 75 replies had been received, the results were tabulated as in the case of the first set. Meantime, the judgments of 41 individuals especially competent to judge merit in English composition writing were secured to use as a check on the others. The examination of the results from this second set showed the necessity of adding two more samples to the set. This was done, making 29 samples in all.

After one or the other of the two sets, to which 21 of the samples were common, had been judged by about 200 individuals, it was decided to make the scale. The first thing necessary was to locate a zero point. This point was to be represented by a sample which possessed absolutely no merit as an English composition. It was chosen on the basis of the judgments of 28 qualified individuals. When the result of these judgments was tabulated, it was found that just one-half of them considered such a point as below sample 580 and one-half as above it, and so sample 580 was taken as the zero point on the scale.

The ten samples chosen for the scale were selected on the principle of equally often noticed differences, which is as follows: Differences that are equally often noticed are equal (unless always or never noticed). Thus, if in a set of samples,  $a$ ,  $b$ ,  $c$ ,  $d$ , etc., it was found that  $a$  was judged better than  $b$ , just as often as  $b$  was judged better than  $c$ , and so on, samples  $a$ ,  $b$ ,  $c$ ,  $d$ , etc., would constitute a scale of equal steps. To put the case more concretely,

if in an essay contest, essay *A* was judged better than essay *B* in 75% of the judgments, essay *B* was judged better than essay *C* in the same number of judgments, and so on, it is readily seen that the differences in quality between essays *A*, *B*, *C*, etc., are equal because the same number of individuals noticed this difference. Similarly, as a result of all the comparisons made of the sample compositions, the result was approximately as follows:

Sample 18 was judged better than sample 0 in 75% of the judgments.

Sample 26 was judged better than sample 18 in 75% of the judgments.

Sample 37 was judged better than sample 26 in 75% of the judgments, and so on for samples 47, 58 and 94.

Thus in samples 18, 26, 37, etc., we have the successive steps of a scale, steps that are equal inasmuch as they represent differences that are equally often noticed.

Why the opinion of 75% of the judges was taken as the unit of value, instead of some other per cent, may probably be better understood if the following case is considered. If, in comparing the ability of two statesmen, say Gladstone and Bismarck, 50% of the judges claim Gladstone to have possessed the greater ability, while 50% claim the same for Bismarck, it may safely be assumed that they possessed about equal ability. If, however, 60% of the judges believe Gladstone to have been the more efficient, the chances are that Gladstone was probably slightly more capable than Bismarck. As the percentage of judgments favoring Gladstone increases, the chances are shown to be greater that Gladstone had the superior ability, and when 100% of the judges believe him to have surpassed Bismarck it may safely be assumed that such was actually the case. Similarly, in the present case if 75% of the judges say that a given sample is better than another given sample, we may be reasonably sure that such is the case.

The value of any English composition may be obtained by placing it alongside the samples in the scale and deciding which it is most nearly like in quality. By having other judges measure it, each being in ignorance of the judgment of the others, or, if this is not practicable, by rating the sample two or three times, a very accurate measure of it may be secured. For example, if the composition seems to be very similar in quality to sample 77, then it is marked 77. If it seems to lie between samples 77 and 83, it should be given a value between 77 and 83, as 79 or 81, according to which sample the specimen more nearly resembles.

## II. HARVARD-NEWTON SCALES

An experiment with the Hillegas Scale in the public schools of Newton, Massachusetts, led the school authorities of that city to believe that it possessed several inherent defects. They maintained that since the scale provides one, and only one, type of composition for each one of the grades, the type of one grade differing entirely from that of the next (that is, grade *A* in the scale is represented by one type of composition, grade *B*, by another, and so on), it was difficult or impossible to compare the work of one type of composition, narration, for example, with that of another type, like description. Moreover, they claimed the sample compositions were not typical of efficient school work. An attempt to remedy these defects resulted in the Harvard-Newton series of scales, the general nature of which will be described before the construction of the scale is discussed in detail. This objective measure is the outcome principally of the coöperation of Ballou, and the teachers of the Boston and Newton public school systems.

It consists of four separate scales to measure the four different forms of composition in the eighth grade; namely, description, narration, argumentation, and exposition.

Each scale in the series is composed of six compositions, actually written by eighth grade pupils; thus each scale possesses the same qualities that it is designed to measure. These sample compositions range by approximately equal steps from the best to the poorest work which is likely to be done in the eighth grade, and each of them has been assigned a letter and a percentage valuation in conformity with the current practice in grading. "A" represents the conventional value of 95%; "B" that of 85%; "C" of 75%; and so on. In this way sample "A" is fairly representative of all compositions whose value would seem to lie somewhere between 90% and 100%; sample "B", of all those whose value would seem to lie between 80% and 90%, and so on. Each sample composition is accompanied with a short description of its merits and defects, and it is compared with the next higher and lower compositions in the scale. These descriptions and comparisons were written by the teachers who helped to make the scale and expected to use it. Without some such guiding material, it is doubtful whether those who use the scale would see the same merits and defects in a composition as those who made the scale, and, unless this was the case, little advantage would be derived from its use. The general nature of the four scales may readily be seen from the one—the description scale—which follows.

## THE COMPLETED DESCRIPTION SCALE

No. 1. "A" GRADE COMPOSITION. VALUE, 94.6%

## A STORM IN A FISHING VILLAGE

It was a cold damp day in November. The sky was a heavy leaden color. In the east a black line stretched across it foretelling the coming of a storm. The houses across the way were dismal shadows, — flat, cold, heartless. A piercing chill penetrated to the bone. The rattle of a grocer's cart or the clatter of a horse's hoofs, seemed cold. The pedestrians were all clothed in black, or else the feeble light made them seem so, and they were cold — everything was cold, cold, cold. An awful loneliness pervaded all.

The black line in the east had grown into a cloud and was coming nearer, nearer, over the sea. Suddenly a gust of wind shook the very foundations of the houses, — another, and then a continuous blowing. The howling was horrible. Great sheets of foam were blown into the streets, — here and there a piece of wreckage hurled itself against a cottage. Fishermen's wives hurried down the narrow streets to the shore, straining their eyes for any sign of a wreck. Old seamen looked at the roaring sea and shook their heads.

By this time the black cloud had engulfed the sky. The day was like night, although it was not yet noon. Boys ran about with torches which were immediately extinguished, and the roaring called to mind the last day at Pompeii.

Rain had begun to descend. At first only drops fell on the hardened faces of old mariners, and on the pale countenances of wives, mingling with the drops already there. But soon great sheets fell, forcing the people indoors, to the poor shelter afforded by the groaning houses.

For about an hour the storm continued thus, then by degrees the wind lessened, though the rain still fell, and the ocean thundered. But soon the rain also slowly stopped and the roaring ceased. The black cloud rolled slowly away, leaving the tardy sun to shine on the drenched town and the great piles of wreckage on the shore. 35

### Merits

This theme ranks high because the writer has a clear picture of the scene and has used words and phrases that bring the details of this picture clearly before the reader. There are good color images in such expressions as *leaden, a black line, great sheets of foam, the day was like night, and the sun shining on the drenched town*. Sound effects are strikingly brought out by such phrases as *the rattle of a grocer's cart, the howling, the wreckage hurled against the cottage, the roaring sea, and the thundering ocean*. The sensation of dreariness and chill is conveyed by the repetition of the word *cold*. The confusion caused by the storm is reflected in the anxious look of the wives of the fishermen. A further human touch is added in the mention of such details as *the extinguished torches carried by the boys and the drops of rain falling upon the hardened faces of the old mariners*. All these enumerations fittingly combine to produce a tone of coldness, desolation, and anxiety. The details are told in their natural sequences. This chronological arrangement has helped the writer to keep safely to his main point and effectively connect the details with each other.

### Defects

The repetition of the word *cold*, while effective in bringing out the sensation, is somewhat artificial. *Loneliness* (line 9), is misspelled; a semicolon should supplant the comma in line 8. Omit the comma in line 6.

### Comparison

The theme is superior to No. 2 in its richness of imagery, its wealth of details, its depth of feeling, its maturity of style (seen in the sentence-structure and the vocabulary), and in its mastery of mechanical forms.

No. 2. "B" GRADE COMPOSITION. VALUE, 83.5%

### GRANDMOTHER

In front of the open fireplace in a large armchair there sits our old Granny. She is old and feeble. Her hair is snow-white and over her head a little white cap is carefully tied. Her face is full of wrinkles and her keen blue eyes sparkle through a pair of glasses which she has on her nose.

She has a shawl thrown over her shoulders and she also wears a thick black skirt. On her feet can be seen a pair of soft slippers which she prizes very much because they were given her for a Christmas present.

As you know Grannies always like to be busy our Granny is busy knitting gloves. Her hands go to and fro. She will keep on working until her knitting is done. Now that it is done she carefully folds her work and packs it into her workbasket. Then she trots upstairs to bed and oh, how lonesome it is when our dear Granny is gone from the room.

### *Merits*

The merits of this composition are: (1) the clear and pleasing impression obtained; (2) the happy choice of details and the logical sequence of their arrangement; (3) the sympathetic treatment of the subject — for example, bits of sentiment seen in the grandmother's attachment to the slippers, and the loneliness felt when she goes to her room; (4) the interesting introductory sentence; and (5) the mechanical accuracy.

### *Defects*

The defects are: (1) the rather monotonous sentence structure, and (2) the childish vocabulary.

### *Comparison*

To justify its place in the scale, note: (1) that in No. 1 there is successfully treated a much more difficult subject; (2) there is a greater power of imagination; and (3) there is a greater variety of sentence structure and a richer vocabulary.



No. 3. "C" GRADE COMPOSITION. VALUE, 76.1%

### A MANSION

As you look across the road you will first see a long private avenue or walk.

It is in the summer, and on each side of this long walk are some beautiful, stately elms. They are hundreds of years old and they have done their duty for as many, 5 years, shading the walk from the noon sun.

Cross the road and you will see if you look up the avenue, a beautiful mansion. It is a colonial house and four large pillars are upholding the roof. A piazza runs along three sides of the house.

10

Near the house is a tennis court where for years the occupants of the mansion have passed many an hour.

Let us enter the mansion. It is a beautiful cool place, although dark. As we enter we see large psalms on each side of the entrance. On the floors are old oriental rugs 15 which have been handed down for generations. In the parlor is a harp, and on the walls are the portraits of the ancestors. In all, it is a beautiful place.

### Merits

The writer of this theme has presented a clear though conventional picture. Although he changes his point of view several times, he has attempted to put his readers into the best positions to see the mansion. The choice of words is fair. Such details as the *stately elms*, the *oriental rugs*, the *harp*, and the *portraits* are well selected. Only one mistake in spelling occurs (line 14).

### Defects

There are, however, too many paragraphs for such a short theme. Constant repetition of the pronoun *you*, and of the words *beautiful* and *mansion* give an impression of monotony and of limited vocabulary. The pupil has evidently a definite place in mind, but has not suggested the spirit of the scene, as has the writer of No. 2.

*Comparison*

The composition deserves its place in the scale above No. 4 because of better sentence structure and more orderly arrangement. It is inferior to No. 2 on account of its somewhat prosaic tone and its constantly changing point of view.

No. 4. "D" GRADE COMPOSITION. VALUE, 66.6%

## THE LAKE AT SUNRISE

In the Mountains of Pennsylvania there is a lake.

On one side of the lake is a boat landing, at which a dozen or more boats are tied up. On this boat landing one may stand and look up the lake, at sunrise, and see  
5 the sun peering up over the top of the mountains and shinning on the water. Then a King Fisher flies down the lake making his cheerful noise, instantly, all the other birds begin to chirp as if their life depended on it.

Looking across the lake one would see numerous wells  
10 and coves backed up by woods from which comes the chirp of the birds. Hearing the explosions of cylinders we look to see where in comes from and find a pumphouse that keeps the lake supplied with water.

Looking down the lake over the dam to the ice house  
15 with the roof sparkling with. On the roof of the house a hawk is sitting adding his clear whistle to noise of other birds.

Looking around to the woods, at our back, with an old oil well in front of them. The birds flying from the woods  
20 in flocks, and far away from the hills comes the sound of the of Italians singing.

*Merits*

The writer has seen and heard concrete details and has re-created his images clearly. He has tried, too, to make his point of view obvious to the reader. His vocabulary is adequate.

*Defects*

As a description the composition fails because there is no unified picture of the lake. The selected details, clear in themselves, tend to distract rather than center the interest. There are numerous mechanical errors: there should be no commas after *lake* or *sunrise* (line 4); *shining* (line 6) is misspelled; there should be a period after *noise* (line 7), and no comma after *instantly* (line 7), which should commence with a capital; *in* (line 12) is not correct; the groups of words in lines 14, 15, and lines 17, 18 do not make sentences; the word *the* is omitted before *noise* (line 16) and the word *are* before *flying* (line 18).

*Comparison*

The theme merits its rank in the scale by superiority in spelling, paragraphing, and maturity of thought. It does not, on the other hand, show equal mastery in the fine details, the discriminating vocabulary, and in the ability to stick to the point. The sentence-sense is faulty.

No. 5. "E" GRADE COMPOSITION. VALUE, 55.4%

*A LIGHT HOUSE*

A description of a light house is quite interesting.

First a light house is generally situated on a mass of rocks in the ocean or on some great lake. And then to get into a light house is a question. Some times you have to climb to the top on a steal ladder, and again you only 5 have to go half way up and you find sort of a steal porch. which is very strong with a door in the side of the light house. On the very top of the light there is generally two or three life boats in case of accidents. In side there is an enormous light which flashes every two minutes and 10 sometimes more often it depends holy on the weather. The man himself has very favorable sleeping quarter and food it is a very lonely life except when you have a man with you. Sometimes they play cards all day long until it is time to fix the lights and then they are very busy. 15

*Merits*

The merits of this theme are: (1) the evident spirit of faithful accuracy; and (2) a successful use of certain simple words, — such as *mass of rocks*, *enormous light*, and *lonely life*.

*Defects*

Many obvious defects warrant its low position in the scale. The pupil was asked to write a description. After announcing his purpose to do this, he writes an exposition, or explanation of lighthouses in general. The first sentence of the theme is worthless, contributing nothing toward the development of the subject. It should be omitted. The paragraph is full of misspelled words and grammatical slips; *steal*, *in side*, *holy*, *some times*, *sleeping quarter*. The most striking weakness of the work is the loose and rambling form of the sentences, indicating indefinite thought. "Run-on" sentences are found in lines 9–13. No attempt has been made to establish a point of view. On this account, and because of a lack of vivid words, the passage is dead and colorless.

*Comparison*

The composition is placed above No. 6 because it contains fewer mechanical errors.

NO. 6. "F" GRADE COMPOSITION. VALUE, 44.9%

## A SCENE ON THE PRAIRIES

Along a large plain in the west with mountains on all sides. The sun was just sinking behind the mountains. Some trappers were on the plain just about to get their supper. They had one tend because there was just three  
5 of them. Beside their tent tripled a little spring. After the three trappers had eating there supper they sat down by the fire because it had growing dark. All of a sudden a bunch of Indain's came riding up. When they came near they fired of their guns and disappered in the dark-  
10 ness and the trappers turned into camp leaving one a the trappers on gaurd.

---

*Merits*

The commendable features of this composition are directness, simplicity, and a logical arrangement of details. The writer passes from the general to the specific in a natural manner. In spite of a change in the point of view in the last two sentences, the paragraph, as a whole, makes a clear picture.

*Defects*

Blunders in grammar and in spelling, lack of sentence-sense, and short, childish sentences make the rating of the composition necessarily very low. Such errors as *tend* for *tent*, *tripled* for *trickled*, *eating* for *eaten*, *growing* for *grown*, and the misspelling of *Indians* indicate either hasty, careless work, or slovenly habits of enunciation.

*Comparison*

Compared with the descriptions of the storm and of grandmother, the short sentences here show immaturity and weakness rather than skill or force. With a large amount of correcting of mechanical details, but with very little revising as a whole, this composition would be superior to No. 5.

The scales and tables in this section are reproduced by the courtesy of Dr. F. W. Ballou.

## EFFECT OF USING THE SCALE

An initial experiment in the use of the description scale was made in Arlington and Boston. Eighth grade teachers and elementary school principals in these two cities graded a set of twenty-five eighth grade compositions secured for this purpose, both without the use of the scale and with it. With the use of the scale the results showed a reduction in the extreme variation of judgments; that is, no two teachers were quite so widely divergent as before. The average variation was also less. But in this matter neither the average nor the extreme variation is the most important consideration. Far more important is the effect which the use of the scale has on the grading of each individual teacher. To ascertain this is obviously a complicated matter, and it requires more time than has been thus far at our disposal. This phase of the problem will be the subject of further investigation.

The compositions used in the scale were selected from a large number written by the eighth grade pupils of Newton as a part of their regular school work. Each pupil was given his choice among several topics of description, narration, exposition, and argumentation, suggested by himself or the teacher, and was required to write a composition of about a page in length. Time for preparation and correction was allowed. Thus, these compositions represented the best unaided writing of the individual children in the eighth grade of that particular city. Then a selection from all these compositions was made by the individual eighth grade teachers. This selection included at least 25% of all the compositions written in a particular class and was made with the view of securing compositions representing all degrees of ability in that class. The compositions were then numerically graded by the eighth grade teacher and the principal, independently. To be sure of securing compositions deserv-

ing the highest grade of merit, namely, "A" or 95%, each school, in addition, sent in from one to three of its "best" compositions in all four types of writing, as judged by the teacher and principal. Twenty-five samples of each one of the four types of composition—description, narration, exposition, and argumentation—seemed a sufficient number from which to select the six compositions to be used in the final construction of each one of the four scales. Twenty-five samples, then, of each type were selected on the basis of the preliminary grading given the compositions by the teachers and principals and on the judgment of Ballou, director of the experiment.

To eliminate any possible influence of handwriting these samples were typewritten and mimeographed. Then one set, consisting of 25 samples of each of the four types of composition, was sent to each of the eighth grade teachers and principals, 25 in all, with instructions (1) to grade each of the compositions independently and (2) to rank each in the order of its merit.

Because of the probability that 95% rather than 100% would represent the highest degree of efficiency in composition writing in the eighth grade, and because it was desirable that each reader should start from the same point in marking the compositions, the teachers were asked to give 95% to the best compositions. Although no lower limit was fixed, 40% was intended to be that limit; for compositions worth less than that were not to be furnished by the schools for the experiment.

As already stated each composition was graded by 25 teachers, and, when the marks came in, five things were noted with regard to each of them:

- (1) Its average mark (found by dividing the sum of all the marks by 25).

- (2) Its median mark (found by ranging all the marks given it in order from the highest to the lowest and taking

the middle one). (This is easier to find than the average and for many purposes it is better.)

(3) The highest mark given it.

(4) The lowest mark given it.

(5) The difference between these two, which is the maximum variation in the marking of these particular compositions.

MARKS GIVEN TO THE TWENTY-FIVE COMPOSITIONS

COMPOSITION NUMBER	HIGHEST GRADE	LOWEST GRADE	MAXIMUM VARIATION	MEAN OR AVER- AGE GRADE	MEDIAN <sup>1</sup> GRADE
1	95	68	27	91.9	83.0
2	90	64	26	80.0	80.0
3	50	30	20	42.7	41.0
4	94	63	31	84.3	85.5
5	78	50	28	61.1	60.0
6	88	50	38	69.4	69.5
7	80	40	40	63.5	65.0
8	95	52	43	82.3	85.0
9	75	40	35	56.1	58.5
10	95	90	5	94.5	95.0
11	65	40	25	49.5	49.5
12	75	42	33	59.9	60.0
13	95	71	24	83.7	85.0
14	76	40	36	55.4	53.5
15	95	80	15	89.6	90.0
16	92	68	24	78.2	78.5
17	93	63	30	81.0	81.5
18	90	60	30	79.9	75.0
19	92	60	32	79.6	80.0
20	92	70	22	82.7	85.0
21	89	54	35	76.1	77.0
22	86	47	39	66.6	66.5
23	74	40	34	55.4	57.5
24	73	30	43	48.9	48.0
25	62	20	42	44.9	45.0

As a check on the results of the gradings, the returns from the rankings were also tabulated and the same items noted as in the case of the grades.

<sup>1</sup> "Median grade" is the grade in the series of grades above which and below which there is an equal number of grades.



After the various items in both grading and ranking had been recorded for each composition, using these data as a basis, it was necessary to choose the compositions best fitted to have a place in the scale. It is obvious that compositions about which there was most agreement in judgment on the part of the teachers, both as to rank and grade — that is, compositions with low maximum variations — were most desirable; furthermore, since it was the intention of the authors of the scale that the six compositions selected should represent 95%, 85%, 75%, 65%, 55% and 45%, respectively, in choosing the compositions for the scale they accordingly selected those whose average and median marks came nearest those requirements.

In short, in constructing the scale there were no fixed requirements set. The compositions selected were those about which there was the least disagreement as to merit and whose marks approximated those desired in the scale. After the six compositions had been selected on this basis, the teachers were asked to point out in a brief paragraph the merits and defects of each of the compositions. These paragraphs were carefully studied and compared by a committee who, acting under expert advice, put the various criticisms into the form shown in the scale already presented.

The method of using this scale is very simple. The composition to be measured is compared directly with those in the appropriate scale — description, narration, etc. — and its value determined in terms of the marks assigned to the sample composition which it most nearly approaches in quality. Thus a descriptive composition is placed alongside the compositions in the description scale, a narrative composition alongside the compositions in the narration scale, etc. If the composition to be measured seems to possess the same qualities as a given composition in the scale — say the composition represent-

ing grade "B" in the description scale — then it is assigned the same value as that composition, namely grade "B" or 83.5%. If its value seems to lie somewhere between two grades on the scale as represented by two compositions, say "A" (94.6%) and "B" (83.5%), the examiner can determine its value as precisely as he pleases according to its apparent distance below the one and above the other.

In spite of the difficulty of comparing a sample of composition writing of one type with a sample of another type, as is necessary in using the Hillegas Scale, in actual practice the Hillegas Scale has on the whole been used to greater advantage than the Harvard-Newton Scale. This has been due chiefly to the fact that the field in which the former may be used — the elementary grades and high school — is not as limited as that of the latter, which is confined to the eighth grade. However, for eighth grade measurements the Harvard-Newton Scale may obviously be used to better advantage.

The teacher may obtain the Hillegas Scale by sending to Teachers College, Columbia University, New York. To recapitulate, all that need be done in using it is to slide the composition to be measured along the scale — as in the case of the handwriting scales — beginning with the sample marked 0, until a sample is reached on the scale to which the specimen to be measured most closely corresponds in quality. As has been said, the former may be of an entirely different type from the latter. The composition to be measured is then given the same value as the one on the scale to which it is most similar in quality. That is, if it appears to be very like the composition marked 77 it is given the value 77. If it seems to be better than composition 77 but not so good as the next composition in the scale, number 83, it is given a value somewhere between 77 and 83 such as 79 or 81.

Teachers of the eighth grade may obtain the Harvard-

Newton Scale by sending to The Harvard University Press, Boston, Mass. In using it, a descriptive composition is measured by comparing it with the sample compositions on the description scale, a narrative composition, by comparing it with samples in the narration scale, etc.

Whichever scale is used, in obtaining the compositions to be measured, the teacher must see first of all that the same amount of time for writing is allowed to all the pupils and, secondly, that the same subject is given to all to write upon. Even in thus making the conditions under which the compositions are obtained as objectively uniform as possible, it is apparent that certain subjective influences, such as interest for example, which cannot be eliminated, are bound to affect the result. Furthermore, within the same class there will be the widest difference in the amount of material written.

While it is evident that in disregarding these two factors the scales are not complete as adequate measures of composition writing, still they are of great value; for by their use the composition work of any grade, school, or system of schools in any part of the country may be compared with that of any other, and the results of different methods of instruction or of other conditions ascertained and utilized. Moreover, there is so intimate a relation between the successful use of oral and written language and intelligence that an objective standard which accurately measures ability in the use of language also measures, to a certain extent, the possession of mental ability in general. In the writing of English composition, whatever its type, children are compelled, or should be compelled, above everything else to make themselves clear, and, by the use of a uniform standard of judgment, the growth of reason itself, from grade to grade, may be followed and subnormal or supernormal children detected. Then, too, the difference shown by the same child in the

various types of composition may give a fair idea of his individuality. Increased knowledge of the various types of pupils with which the school has to deal will naturally lead to greater variety in teaching and correspondingly better results with the children. Any such educational progress, however, will come not as an expression of mere opinion, but as the result of scientifically determined educational facts obtained by the use of objective standards. The more scientific, yet comprehensible, are our methods of investigation, the more valuable will be their results.

### EXERCISES

1. In what way may these scales be utilized to secure a very accurate judgment of the merit of a given composition?
2. What relation seems to exist between ability in composition writing and ability in other subjects in the curriculum? Between ability in composition writing and general intelligence?
3. Procure twenty compositions from various grades and get five teachers to mark them on a percentage basis. What do the results show regarding the reliability of such measures?
4. How would the ratings given by five teachers to twenty compositions of varying merit test the reliability of the Hillegas Scale?
5. Suppose the Composition Scale revealed a great difference in the same child in the various types of composition writing, of what value would this be to the teacher?
6. Obtain forty specimens of English composition from the various grades. Grade these on the Hillegas Scale. Allow one month to elapse and grade again. What do the results show?
7. In what type of composition writing do you think a child should be most proficient?
8. Suppose a teacher discovered by the use of the scales that the pupils on the whole showed far greater efficiency in one type of composition writing than in another, what should be the conclusion?
9. How would you modify the standard for composition writing for your particular grade? Why?
10. What modifications would you make in it if your pupils came from a foreign neighborhood?

## CHAPTER VII

### COMPLETION TEST LANGUAGE SCALES—TRABUE

Suppose we consider an incomplete sentence such as the following: "The . . . rises . . . the morning and . . . at night," where three words are omitted, the place of each word being filled by a dotted line; it is a simple matter for any one who is acquainted with the English language to insert a word in each of these three blank spaces, which will cause the sentence to make sense. In the above example, these words are "sun," "in," and "sets," making the sentence read: "The sun rises in the morning and sets at night." The completion of sentences of this kind, while not actually testing ability in English composition, demands an ability very closely related to what is usually called "language ability"; at any rate, it involves a power to read and think about printed words which has great educational significance.

From the nature of this test it is obvious that we may have sentences for completion of all degrees of difficulty. While a sentence such as, "The sky . . . blue," requires next to no ability in English language, a sentence such as the following; "To . . . friends is always . . . the . . . it takes," is of sufficient difficulty to test the ability of a college student. If, therefore, we could select a series of incomplete sentences increasing in difficulty from the first to the last, with this as a scale, we should be in a position to measure the language ability of any individual or group. This could be accomplished by allowing a certain specified time in which to complete as

many of the sentences as possible. To construct such a scale for the measurement of language ability of this type was the object of the study made by Trabue.

A large number of incomplete sentences were constructed. After a preliminary trial fifty-six of these sentences were selected and their relative difficulty determined by administering them, under standard conditions, to several thousand children and young people in various school systems. The detailed scheme by which each sentence was marked will be described later, but the general method was to give a score of 2 for a perfect completion, a score of 1 for an almost but not quite perfect completion, and a score of 0 for a failure to attempt or for an imperfect completion.

By determining the different scores made on the sentences in the various grades, it was possible to calculate the relative difficulty of each of these sentences. Thus, two sentences were considered of equal difficulty when they were completed by the same percentage of individuals tested. The greater the difference of percentage attained in completing two sentences, the greater was the difference in the difficulties of the sentences. It is impossible to enter into the details of these calculations, but the method employed was essentially the same as that described in the construction of the Buckingham Spelling Scale.

Knowing the difficulty of these original sentences, Trabue constructed eight short scales. The following are some of the reasons for the use of several short scales:

- (1) A short scale takes less time to administer and score;
- (2) a measure of ability is more reliable when taken on two separate occasions than when taken at one time;
- (3) a number of scales of equal difficulty admit of a class being tested from time to time, the use of different scales being necessary to eliminate the factor of memory.

Two scales, called by the author B and C, are here shown; in the study six similar scales are also given.

### Language Scale B

*Write only one word on each blank. Time limit, seven minutes.*

*Name*.....

1. We like good boys.....girls.
6. The.....is barking at the cat.
8. The stars and the.....will shine tonight.
22. Time.....often more valuable.....money.
23. The poor baby.....as if it were.....sick.
31. She.....if she will.
35. Brothers and sisters.....always.....to help.....other  
and should.....quarrel.
38. ....weather usually.....a good effect.....one's  
spirits.
48. It is very annoying to.....tooth-ache, .....  
often comes at the most.....time imaginable.
54. To.....friends is always.....the.....it takes.

### Language Scale C

*Write only one word on each blank. Time limit, seven minutes.*

*Name*.....

2. The sky.....blue.
5. Men.....older than boys.
12. Good boys.....kind.....their sisters."
19. The girl fell and.....her head.
24. The.....rises.....the morning and.....at night.
30. The boy who.....hard.....do well.
37. Men.....more.....to do heavy work.....women.
44. The sun is so.....that one can not.....  
directly.....causing great discomfort to the eyes.
53. The knowledge of.....use fire is.....of.....  
important things known by.....but unknown.....  
animals.
56. One ought to.....great care to.....the right.....of  
....., for one who.....bad habits.....it.....  
to get away from them.

The scales in this section are reproduced by the courtesy of Dr. M. R. Trabue.

Each of these scales consists of ten steps or sentences, the intervals between the various sentences being approximately equal; that is, sentence 6 is as much more difficult than sentence 1, as sentence 8 is more difficult than sentence 6, and so on. It should, however, be noted that Scale C is, on the whole, a little harder than Scale B, Sentence 2 in Scale C is a little more difficult than sentence 1 in Scale B, and sentence 5 in Scale C is a little more difficult than sentence 6 in Scale B. The same is true throughout the series.

### Directions for Administering the Test

The scales which have been described may be purchased in any quantity from the Bureau of Publications, Teachers College, New York. It should be noted that these standard blanks must be used if the results obtained are to be used for comparative purposes. When the test is given to a third or lower grade, it is necessary to give a little preliminary training, using a practice sheet, which can be secured with the regular tests. In the fourth grade and above, the following oral explanation should be made before distributing any papers:

This sheet contains some incomplete sentences, which form a scale. This scale is to measure how carefully and rapidly you can think, and especially how good you are in your language work.

You are to write one word on each blank, in each case selecting the word which makes the most sensible statement.

You may have just seven minutes in which to sign your name at the top of the page and write the words that are missing. The papers will be passed to you with the face downward. Do not turn them over until we are all ready. After the signal is given to start, remember that you are to write just *one* word on each blank and that your score depends on the number of perfect sentences you have at the end of seven minutes.

If there are no questions, the papers may then be distributed, taking care that no child looks at the printed



side until there is a paper upon the desk of each child and the following additional instructions have been given :

After you have been working seven minutes, I shall say, "The time is up. All stop writing!" You will all please stop at once and lay aside your pens (or pencils). Now if you are all ready, you may turn your papers, sign your names and fill the blanks.

Take note of the exact time at which the signal to start was given, allow exactly seven minutes, and give the command to stop writing. Collect all papers at once. It is very important that exactly seven minutes be allowed. A stop watch is the most satisfactory means of keeping the time on a test of this sort. Grade each paper according to the general scheme about to be described, and make a record of the total number of points made by each pupil, in order that the amount of progress of each individual may be determined when this scale is used for a second time, or when another scale is employed. Then arrange the scores in ascending order and find out the median score; namely, that point above and below which there are an equal number of scores. This median value may then be compared with the medians obtained by other classes.

### General Scheme of Scoring

The following general scheme has been the basis upon which the more detailed judgments have been based :

#### *Score 2*

A score of 2 points is to be given each sentence completed perfectly. Errors in spelling, capitalization, and punctuation should not be allowed to affect the score.

#### *Score 1*

A score of 1 is to be given each sentence completed with only a slight imperfection. A poorly chosen word or a common grammatical error, which makes the sentence less than perfect and yet leaves it with reasonably good sense, should serve to reduce the score from 2 to 1.

*Score 0*

A score of 0 is to be given if the sentence as completed has its sense or construction badly distorted. A sentence must have reasonably good meaning and express a sentiment which might honestly be held by an intelligent person in order to receive a higher credit than zero.

It is apparent that the above method of scoring leaves more than is desirable to the judgment of the person who is rating the sentence. This subjective element in the marking is much reduced, however, by a careful consideration of the examples given by the author of what in his opinion constitutes a sentence worth the score 2, 1, and 0, respectively. For illustration take the sentence :

30. The boy who . . . . . hard . . . . . do well.

*Score 2*

works, tries, studies, thinks, . . . . . will,

✓ *Score 1*

tries . . . . can, may, does, shall, should, could, must,

worked, tried, . . . . did, will, can, . . .

plays, hits, work, . . . . will,

*Score 0*

tries . . . . sometimes, surely, often,

did . . . . work did, does . . . . work, work . . . . did,

All the other sentences are treated similarly. The reader is referred to the original study for these completions, as they are too bulky to warrant introduction here.

It will be noticed that the score is given for the whole sentence; that is, in those cases where more than one blank appears, the mark is not given for each single completion but for the whole sentence.

To summarize: All that is necessary to test a class in the type of language ability measured by these scales, is to procure the standard blanks from the publishers. Follow carefully the directions for the administration of the test. Score the tests according to the scheme outlined. Determine the score below which and above which

there are an equal number of pupil's records, and then compare this median value with previous records, if such have been obtained; if not these first results will establish tentative standards.

### EXERCISES

1. How could you rank five completion tests, of your own construction, according to their difficulty? What is the test of a suitable sentence for a particular group?

2. What is the advantage of having the difficulty of the sentences in the scale rise by equal increments? What would happen if three of the sentences were of the same difficulty?

3. How would you use completion tests for determining whether the pupils had read a certain assignment of history or geography?

4. To what extent do these completion tests measure a valuable language ability? How does this type of ability compare with ability in English Composition in your class?

5. How could the idea of the completion sentence be employed to measure ability in a foreign language?

6. Can you reasonably expect the same standard of work, in this test, from schools in a foreign district, and in an English-speaking district? How could a school of the first type establish its own standards?

7. How would you determine the standard of your own class in this test, as compared with other classes of the same grade?

8. State how you would compare the standing of your own school with that of another school? What conditions would have to be fulfilled to make this comparison justifiable?

9. Are completion sentences merely a test, or could they be used with advantage as an exercise to increase thought in language lessons?

10. Suppose a grade fell notably below its average of the last few years, what steps would you take to meet this decline?

## CHAPTER VIII

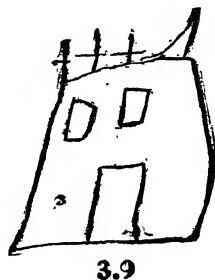
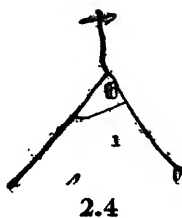
### DRAWING SCALE

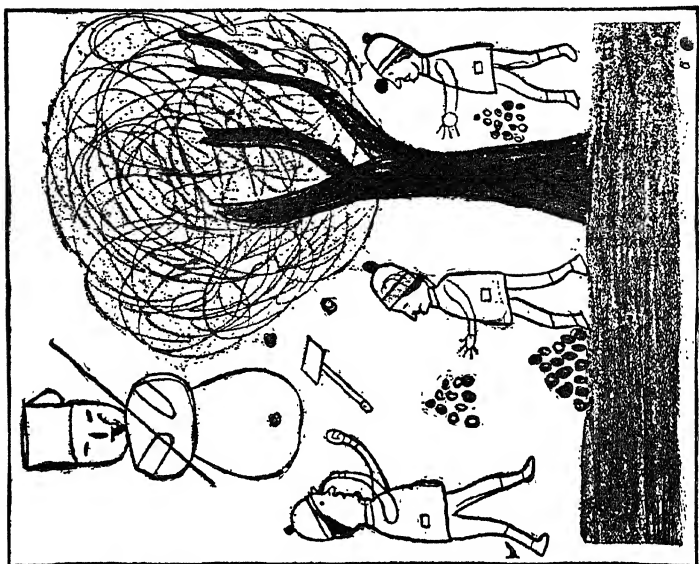
#### THORNDIKE DRAWING SCALE

THE measurement of improvement and efficiency in a subject such as drawing is beset with great difficulties. It is reasonable to suppose that in art the judgment of excellence depends on the individual teacher, to a greater extent than in most of the other subjects in the school course. In spite of this supposition, Thorndike in 1913 presented a scale which, though merely tentative, yet limits to a great degree the possible differences of individual opinion in estimating drawing ability. Its method of derivation is very similar to that employed in the scale for the measurement of English composition. From 45 carefully selected drawings from Kerschensteiner's "*Die Entwicklung der zeichnerischen Begabung*," a more limited selection of 14 drawings was made. These, together with a drawing from another source, constituted the 15 samples. These samples were then submitted to artists, teachers, and students of education and psychology, with the request that they be ranked in the order of merit: that is, that No. 1 be assigned to the drawing which, in the opinion of the judges, is the best; No. 2, to the drawing that is the next best, etc.; No. 15 being assigned to the very worst drawing. It was stated quite clearly that no allowance should be made for the apparent age or training of those who had made the drawings, but that the drawings should all be judged by the standard of their intrinsic merit. In all, 376 ratings or rankings of the 15 drawings were obtained, 60 of which were from

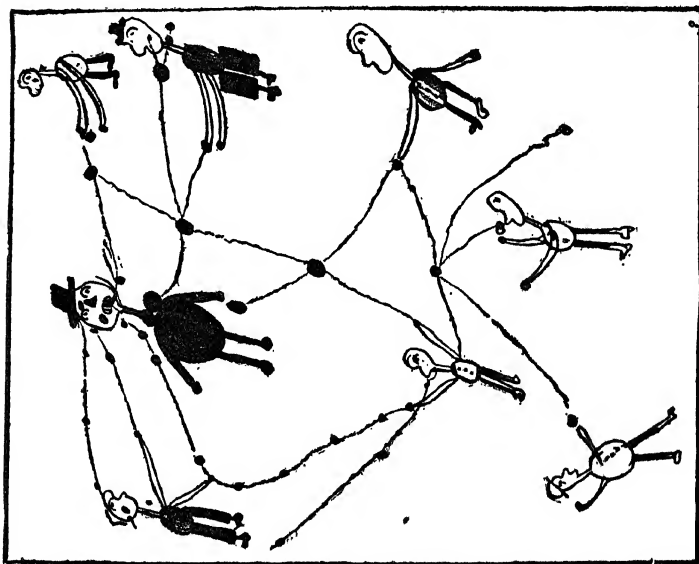
# A Scale for the Merit of Drawings by Pupils 8 to 15 Years Old

The numbers give the merit of the drawing as judged by 400 artists,  
teachers of drawing and men expert in education in general

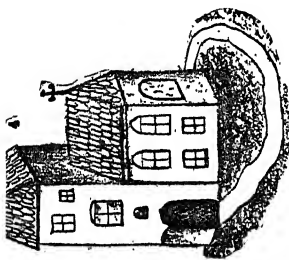




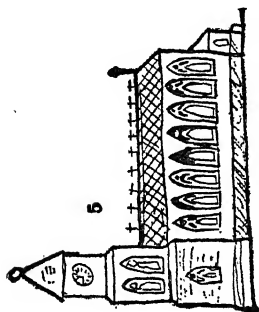
10.5



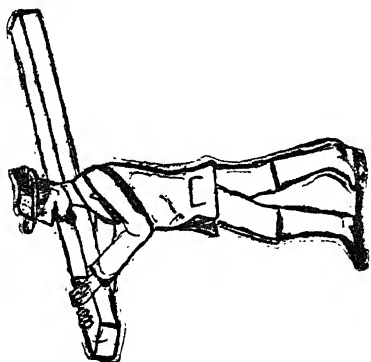
8.6



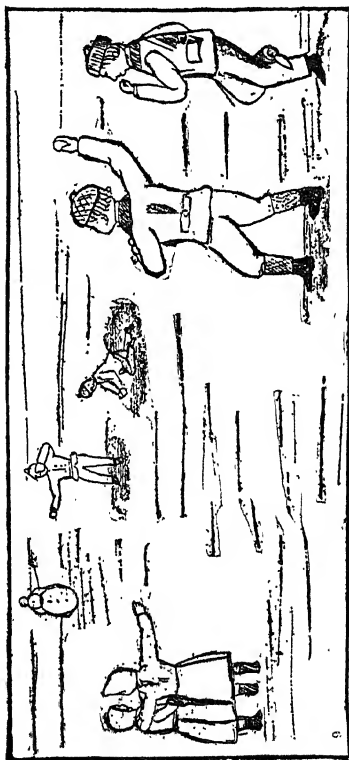
12.6



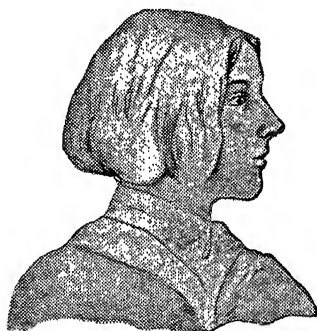
11.8



13.5



14.4

**16.0****17.0**



artists who had sufficient merit to be included in "Who's Who in America."

Suppose the drawings be called *a, b, c, d, e, f, g, h, i, j, k, l, m, n*, and *o*. These fifteen drawings were so chosen that they proceeded step by step from a drawing of almost zero merit, to a drawing of such a high order, that only one child out of five thousand under fifteen years of age was able to produce work of that degree of excellence. When the data which came from all the judges were collected, an idea was obtained of the relative merit of each of the samples of drawing. Thus, suppose it was desired to compare sample *b* with sample *a*, and it was known that 95% of the judges rated *b* as having more merit than *a*, while 85% rated *c* as having more merit than *b*. From general considerations, it was safe to assume that the difference in quality between sample *b* and sample *a* was greater than the difference in quality between sample *b* and sample *c*. This can be seen at once if we consider what it means when 100% of the judges rank one sample greater than another sample; for, in this case, the superior sample is so much better than the inferior sample, that not one judge in a hundred thinks it inferior. Thus, if we compare the plays of Marlowe and Shakespeare by this method, there is such a great difference in quality, that 100% of competent judges would think Shakespeare superior to Marlowe.

Let us consider for one moment what is implied by the statement that 50% of the judges ranked specimen *X* as better than specimen *Y*; in this case, as many judges thought *Y* was better than *X* as thought *X* was better than *Y*. Under these conditions, if the judges are competent and sufficiently numerous, we are justified in assuming that *X* is equal to *Y* in merit. Thus, if 100 people were to compare the merits of two novels, such as "Silas Marner" and "Scenes from Clerical Life," and it was found that 50% of the judges thought "Silas Marner"

was superior and 50% thought "Scenes from Clerical Life" was superior, we should be justified in assuming that the two novels were of approximately equal merit. To summarize, when 100% of the judgments rank *X* as superior to *Y*, then *X* is in all probability very far removed in merit from *Y*; whereas, when 50% of the judgments rank *X* as superior to *Y*, then *X* and *Y* are approximately equal in merit. The results of the rating of the drawings by 187 judges are shown below in the table.

#### RATINGS OF DRAWINGS

94.85% of the judges rated <i>b</i> as better than <i>a</i> .									
84.5	"	"	"	"	<i>c</i>	"	"	"	<i>b</i> .
88.45	"	"	"	"	<i>d</i>	"	"	"	<i>c</i> .
69.5	"	"	"	"	<i>e</i>	"	"	"	<i>d</i> .
82.55	"	"	"	"	<i>f</i>	"	"	"	<i>e</i> .
69.7	"	"	"	"	<i>g</i>	"	"	"	<i>f</i> .
89.4	"	"	"	"	<i>h</i>	"	"	"	<i>g</i> .
81.75	"	"	"	"	<i>i</i>	"	"	"	<i>h</i> .
70.	"	"	"	"	<i>j</i>	"	"	"	<i>i</i> .
73.35	"	"	"	"	<i>k</i>	"	"	"	<i>j</i> .
72.5	"	"	"	"	<i>l</i>	"	"	"	<i>k</i> .
86.5	"	"	"	"	<i>m</i>	"	"	"	<i>l</i> .
74.2	"	"	"	"	<i>n</i>	"	"	"	<i>m</i> .

By simple, but laborious statistical treatment, which it is unnecessary to discuss here, based on the two facts given above, it is possible to arrange the various samples in an order of merit, and to assign to each a numerical value which is the result not of a single judgment but of the combined estimates of many experts. That is, if we assign zero merit to the first picture, which is supposed to be a picture of a man, then by an analysis of the table just given it can be shown that the second figure, which is intended to be a house, has 2.4 degrees of merit. On the same scale, Figure 3, which is also supposed to be a house, has 3.9 degrees of merit, and

so on, until we reach the last three samples, which have 14.4, 16, and 17 degrees of merit, respectively. A scale so constructed enables us to measure skill and improvement in drawing by methods which are largely objective.

In the matter of assigning actual values to the drawings, care must be taken not to assume that the degree of improvement, say from a sample which ranks 6 to a sample ranking 10, is equal to that from a sample ranking 12 to one ranking 16, in the sense that a rise in a temperature scale from 6° to 10° is equal to a rise from 12° to 16°. In the case of the scale for measuring drawing, this is true in a very limited sense only, but the scale can be used with a maximum return without an understanding of these statistical considerations. In other words, when a teacher says that the average ability of a class according to the Thorndike Drawing Scale is 13.5, it conveys a reasonably definite idea to any other person who is acquainted with that scale. For all practical purposes the samples constituting the scale, as used by the average teacher, might have been lettered instead of having numerical values attached.

When it is desired to measure the ability of a class by using this scale, all that is necessary is to choose a certain model or subject and allow a measured time for the drawing; this time should be varied according to the nature of the subject. The subject and the time allowed should be noted very carefully, so that when the test is given again, all these external conditions may be the same. When the drawings are collected each one is measured by being placed alongside the scale, and its position estimated by the teacher, or by several teachers. If it appears to lie between two points of the scale, an intermediate value may be given.

This still leaves a considerable amount to the individual judgment. In other words, the scale is not by any means

purely objective, for equally competent persons would fail to assign the same degree of merit to the same drawing. This factor of personal opinion can however be curtailed by having several individuals measure the drawing by the scale and taking the average of their judgments. The drawing scale, like any other scale in its beginnings, is very incomplete, since it still remains to work out scales for all the various types of drawings taught in the schools. But any scale is better than no scale at all, and continued use of a drawing scale by teachers will standardize judgments and encourage quantitative thinking, even in this study which at present is so dependent on personal opinion.

#### EXERCISES

1. Take 30 specimens of drawing, distributed through the grades, and mark them according to your usual method; let one month elapse and grade them again. What do the results show?
2. Repeat the above experiment, with the exception that the grading is done by means of the Thorndike Scale. How do the two ratings differ? Compare with the results of the previous experiment.
3. Why is it necessary to take careful note of the time allowed for the test? Why must this be the same when the test is repeated, if the grading is to be used to measure improvement?
4. Why cannot we divide children into two classes — "good drawers" and "bad drawers"?
5. How would you proceed to establish norms or standards for drawing ability in the various grades in your school?
6. On the analogy of the Harvard-Newton Scale, how would you propose to construct a better method of measuring drawing ability?
7. Take 5 specimens of drawing from each of the grades; have these rated on the scale by 5 different individuals. How would these results give you an indication of the reliability of the scale?

## CHAPTER IX

### THE APPLICATION OF THE SCALES IN THE SCHOOLS

**Objective Scales in Other Subjects.** Scales for the measurement of other school products have yet to be evolved; the subject is still in its beginnings. In addition to those scales already described, attempts have been made to measure objectively mechanical constructive ability and ability in the translation of Latin, while scales are in process of formation for the measurement of ability in several of the modern languages, in algebra and geometry, and in some of the natural sciences. One of the authors is at present conducting an experiment, extending over two years, the results of which will standardize the rate of improvement in typewriting using the touch method.

A point of interest arises as to whether scales can be worked out for informational subjects such as history, geography, etc.; for at once we have to face the great difficulty that in subjects such as these we have to measure knowledge of facts or content rather than skill or method. The previous study of the writing scale has little effect on a child's proficiency in writing, but the study of a scale for the measurement of content or facts in history, prior to the examination, renders that scale valueless as a test. For the particular facts can be learned, and the knowledge of these will not indicate any general knowledge of the whole field. This means that in measuring efficiency in certain subjects, we may have to resort to analysis, and use one objective standardized test to measure method, and another more or less sub-

jective test to measure content. Whether it will ever be possible to use a universal and unchanging scale for content values remains doubtful. If a very large number of content questions, sufficiently wide to cover the field, could be standardized as regards difficulty, there is no reason why a purely objective scale, consisting of a few of these questions selected at random, should not be employed.

The adoption of these objective scales for the measurement of school products is bound to establish a scientific attitude in the schools, which will energize and direct the work of the teachers and raise the administrator's task from the realm of mere opinion to the level of scientific judgment.

**Standardization of the Objective Scales.** The standardization of these universal tests will involve a considerable amount of work if accurate and complete norms are to be established. In some cases it may be advisable to have the test standardized not only as regards the product of each grade, but also with reference to age. For example, in handwriting — a distinctly motor function — it may be well to know the quality of work expected at a certain age, as well as in a certain grade. A pupil may be held back in a grade because of failure in arithmetic and reading and so become over age. Under these conditions a motor function such as handwriting may continue to improve normally, so that even though the child is in a low grade, we may expect of him the normal standard product of his age in that subject.

As these tests come into common use in the classroom, the interest of the individual teacher will cease to be confined merely to the average of the particular grade, and attention will more and more be directed to deviations from that average which may normally be expected. In fact one of the great services of these tests is to reveal the great individual differences in ability that exist even

in the same class. A teacher of Grade V will not only be interested in knowing that the average achievement of the class in a particular test, say in the Courtis Test, Fundamentals 7, should be 9.0, but will also appreciate the advantage of knowing how the class groups itself around this average, what are the extreme deviations indicating the lowest and highest type of work in the class. In fact, as will be seen, these scales may be used for a variety of purposes by a teacher who is really interested in the work of the individuals of the class.

**The Relation of the Objective Scales to Continuous School Records.** The application of statistics to education is not really a new idea; in certain realms of administration, such as attendance, per capita cost, etc., such measurements have always been made. What is claimed is that this method should extend to all possible phases of classroom work. The ordinary examination fails to do this; the questions are arbitrary, they are not weighted according to their relative values, there are no objective standards of accomplishment. Until units of mass, length, and time were invented upon which all agreed, it was impossible to express weight, dimensions, and time in terms which would convey the same meaning to all. This was, to a great extent, the situation in education ten years ago. But the time is not far distant when, in many of the essential subjects, the progress of every pupil who enters school will be determined by objective methods. Thus, in a particular function such as writing, we shall measure the ability of the pupil every six months from the time he enters until he leaves school. The same will be true of his ability in the other subjects which the school considers to be of importance and which admit of being measured by universal standards. The enlightened school system will have the progress of every child kept on a chart, a rough sample of which is given on the following page.





In this way it will be a simple matter to determine the exact point at which the pupil failed to advance at the normal rate in any particular line of study. The teacher will be able to see whether failure was confined to a particular subject or whether it also took place in other subjects in the curriculum. If it is found that the child has failed in but one subject and not in the others, then we must assume either that the child was abnormal in that subject or that the method of instruction in that particular branch was not up to the usual standard. If, in addition to this, it is found that the majority of pupils under a particular teacher have failed to advance in this subject alone and not in others, then there is every reason to suppose that it was not the fault of the class but rather the fault of the teacher in failing to give attention to the subject or in using some method which could not produce the average rate of progress. Again, in the case of a particular pupil, it may be found that the failure to progress was not confined to one subject, but that it extended to all subjects. In this case further inquiries must be made. It may have been a matter of arrested mental development, or it may have been due to physical causes or to social conditions which did not admit of the child's spending sufficient time in school.

Such a chart as we have shown can easily be passed on from school to school as the child goes from one neighborhood to another, or it can be passed from school system to school system or from country to country, for the very essence of these universal scales is that they are independent of place and time. School systems, under these conditions, will keep track of every pupil from the time he enters to the time he leaves. In other words, the administrator will cease to deal with mere groups of children and will deal with the individual child.

**Application of the Objective Scales to the Question of Promotion.** Such methods of measurement will bring to

the question of promotion a definiteness which is sorely needed. It is too well known that in many school systems a high percentage promotion does not mean a high standard of work, but rather a lowering of that standard to enable the requisite number of children to pass. As a result, pupils are often found in the higher grades who are totally unable to profit by the relatively advanced instruction given. As long as the present loose methods of measuring school achievements are in vogue, such a state of affairs is inevitable; under the new system a radical change will be possible, for with certain exceptions the presence of a child in a particular grade must mean that he has passed certain points on the scales which measure the various school abilities. If these points have not been reached, then the pupil will not be promoted, for he will be unable to profit by the instruction given.

A teacher will be able to measure the abilities of pupils when they are received in September, and if promotion has taken place in spite of bad previous records, he will at least know of this, and, by pointing to their records, will be able to free himself from criticism on account of their ultimate failure. The position of the efficient and conscientious teacher will be established, not on the insecure basis of the opinion of an often prejudiced supervisor, but on the basis of the actual work of the pupils judged by impartial standards.

**Application of Objective Scales to Vocational Guidance.** Such a chart of improvement will be of great service when the pupil on leaving school requires vocational guidance. The employer will state the requirements of his work in the different school subjects, while the vocational guidance expert, by consulting the chart, can determine the extent to which the pupil measures up to these requirements.

**The Objective Scale as Limiting the Amount of Improvement Necessary.** Again, it is true that in many subjects

only a certain degree of efficiency is demanded by the world. For example, there is no object in being able to write better than is required for reasonable grace and legibility. The handwriting of some children shows a wasted youth! If time is spent beyond a certain point, it is relatively wasted. Yet what guarantee have we that when children reach this point they will no longer be given writing lessons? Under the present subjective system of measurement such a guarantee is impossible, and, if given, is meaningless. When the objective scale is used for measuring handwriting, the matter is perfectly simple; for the child knows that when he reaches a certain point on the scale, provided he keeps up to that point, all formal writing lessons will cease.

**Application of the Objective Scales in Rural Schools.** These scales will find ready application in the rural schools, where the teacher is unable to form correct estimates of the work because small classes do not afford a basis for judgment. With the new methods which these scales introduce, the isolated child in the rural school can be compared with, and in a sense can compete with, children of like age in the city system. In fact, at present one of the authors is comparing, by means of these universal standards, the work of 100 rural school children of a given age, with a random sampling of 100 city school children of the same age. In a sense the results of this experiment will be as definite as measurements made of the pupils' height and weight by means of the foot-rule and the weighing machine.

**The Scales as Revealing the Success and Failure of School Methods.** The purpose of these scales, in fact of the whole subject of educational measurements, is not, like the ordinary examination, to test merely the efficiency of the individual teacher or pupil, but rather to test the efficiency of the teaching process itself. The individuals are examined in many cases, not because of our interest

in them as individuals, but because their work will reveal whether the method which is being used in their instruction is sound. Many of the failures in our schools are due, not to unavoidable inefficiency on the part of the teachers, but rather to lack of knowledge on their part that their efforts are failing to produce the desired results. Were the teachers themselves aware that they were failing, they would certainly attempt to alter their methods. It is lack of definite knowledge of what the pupils are accomplishing, and not incompetence or indifference, which prevents a better adaptation of method to product desired.

For this reason teachers should be willing and eager to submit their work to an impersonal standard, not so that it may be praised or condemned, but so that they themselves may know whether their methods are producing as good results as may reasonably be expected. Teachers should have a more exact knowledge than they have had in the past of those processes which are going on in their pupils, for it is the changes which occur during the school period that must be measured. Over these changes we have more or less direct control; the test of life is too remote. The application of these objective scales enables the teacher to know what is happening, not in terms of mere empty formulæ which unfortunately have become associated with the word "method," but rather in terms of what the pupils can actually do as a result of the instruction given them.

Scientific measurement in education will narrow the limits of the wasteful trial and error method which is always incident to the teaching process, however conscientious the teachers may be. It will also do another great service, for it is undeniable that, by means of these scales, the complacency of a small section of teachers can be disturbed by actually showing them their failure in black and white. The greatest check on inefficiency in

any system is the knowledge that the work of each teacher and the work of each school can be compared with the work of other teachers and the work of other schools. A school which is confronted with indisputable evidence of its shortcomings is in a position to investigate causes, and if necessary to trace them to individuals; such procedure is always the forerunner of progress.

### EXERCISES

1. What would be the chief difficulties in constructing a scale for the measurement of knowledge of American history in the eighth grade?

2. How would you prove to an outsider that there are great individual differences in ability, even in the same class? How should a knowledge of these individual differences affect (a) the amount of matter taught; (b) the method of instruction?

3. What are the chief advantages of continuous school records? Draw up a table, and outline the methods which could be used for recording a child's progress in the fundamental studies, from the time he enters to the time he leaves school.

4. Upon what factors should promotion depend? Have we any right to promote a pupil if he is not up to certain minimum standards? How do the standard scales help to determine these minimum standards? How does too lax a promotion system disorganize the work in the higher grades?

5. Why is the present system of marking in your school an insufficient guide to the quality of the work which is being done?

6. Should all children give the same time to all studies? In what way will the use of these standard tests enable us to allow the individual child to distribute his time in a more advantageous manner?

7. How is a rural school teacher handicapped in judging the work of her pupils? Show how the scales help in this respect.

8. A superintendent of a city school system cannot decide between two proposed methods of teaching handwriting. Describe a plan whereby, in a few years, he could decide which method was the better? How have such questions been decided in the past?

9. Why is it better to measure the success of a year's work by the improvement of the pupils during that period, than by the final scores in a test at the end of the year? If you were the principal of a school, outline the methods you would employ to measure such improvement.

## CHAPTER X

### DANGERS INCIDENTAL TO THE USE OF THESE SCALES

AT a time when all available pressure should be brought to bear on school systems to introduce objective measurement into the ordinary routine of the school, it seems hardly the occasion to criticise the scales. However, a word of caution may not be out of place as to the dangers which may arise from their application, since their improper use will perhaps prejudice those who make the first attempts at this type of measurement.

**Difficulty of Comparing Methods of Teaching.** It has already been stated that one of the great functions of the scales is to compare the various methods of instruction employed in the teaching of a subject. Great care, however, will have to be taken to prevent mistakes in comparing the relative values of such methods when used in different schools or systems. To know that the work in a particular subject is better in one school than in another is not sufficient to justify the judgment that the method used in the one school is superior to that in the other. In such a comparison several secondary causes must also be considered before any statement is made concerning the relative efficiency of the methods: (1) time allowed in the different schools; (2) personality of the teacher; (3) the type of neighborhood as determining the type of pupil. It will be only by the most careful experimentation, where attention is paid to these points and to many others of less importance, that anything like a scientific application of the scales to the question of the values of

methods will be obtained. The whole subject is full of danger, and many fallacies will have to be avoided. At the present time scientific attention is being directed rather to the construction and use of scales for particular groups than to comparison of procedure values; but such comparison will be possible later, when every school system employs a competent statistician and experimenter capable of conducting genuinely scientific comparative experiments. In short, we must not strive to compare groups that are not alike or hold up standards without due consideration of social conditions. Mere statistics can never dictate final standards of achievement; a standard set up may be too high for one school and not high enough for another. Each school, after working with these scales for some time, can establish standards of its own; but there is always the danger that a standard may be set up which falls short of what should be done. In fact, the unwise use of standards, in this respect, may confirm the school in lax processes.

**Failure of Scales, from the Fact That They Measure Complex Abilities, to Reveal the Point of Weakness in Method.** While these scales will do much to quicken methods used in the schools, it may be well to mention another point which is apt to be overlooked by some who employ such measurements. Thus, a scale may show that the method which has been used is imperfect in that it has failed to produce the desired product; but it does not directly analyze the particular fault. The scales do not tell you what to do, but rather they tell you where you are. A teacher may be conscious that he has failed, but unable, in spite of great efforts, to find out the exact factor responsible for this failure. In much the same way a physician after examination may make the announcement that the organic processes are wrong, but at the same time be totally unable to attribute the cause. Although the present scales, because they measure such complex

activities, do not reveal the exact point at which a teacher may have failed, yet we see in the Courtis Test the beginnings of an attempt to measure the details of what many have considered to be a single process, namely, "arithmetic ability." When more analytical scales have been worked out in other subjects, it will be possible to go into detail and tell the teacher at just what point or points he failed, these small failures accounting for the failure in the wider test. The idea might also be applied to the testing of English composition. As things are now, it is possible merely to tell a teacher that the class has failed to produce as good English composition, as measured on the Hillegas Scale, as might be expected. We are not in a position to say what details are responsible for the failure. But suppose at a later time scales should be used to test (1) punctuation, (2) extent of vocabulary, (3) choice of vocabulary, (4) power of summarization, etc.; then that which we now attribute perforce to general weakness, we shall then assign to weakness in one or more of these factors which can be corrected by special practice. In this way we shall narrow down the limits within which the teaching process can fail without even a knowledge on the part of the teacher that it is failing.

**What the Scales Do Not Measure.** Another objection which may be urged against the scales is that they fail to take into account such factors as interest in the process of learning, the eagerness with which pupils will continue a particular study after pressure is removed, etc. The scale also takes no direct account of the method by which the product is obtained; it does not tell the experimenter whether these results were secured by easy work or by undue pressure on the part of the teacher. The reply is that it is only the objectors who have ever assumed that the scales do measure these things. To illustrate, in an automobile reliability test, the measurement of speed does not tell us concerning the internal mechanism of the



engine; other tests must be used to measure this factor. But if a machine keeps up a high speed for a long period, then as a rule the internal factors cannot be much out of gear. In a precisely similar manner, if a class steadily keeps up its improvement on a particular scale, then it is feasible to assume that the internal factors are not seriously wrong. In the end, bad psychological methods such as undue driving (which is little to be feared in modern education), will yield poor objective results. The scales, however, must not be attacked because they fail in many cases to measure what no competent individual has ever claimed they do measure.

The use of scales also brings with it the danger that the teacher may sacrifice everything in the classroom to the production of work which can be measured objectively, and, as already pointed out, the scales may fail to give sound relative values to different elements involved in that work. To make this point clearer, let us consider for one moment a scale for the measurement of the child's ability to add simple numbers, such as was described in the Courtis test. If the norms insist upon speed, then the teacher will work for speed; if the norm is one for accuracy, then the teacher will work for accuracy; and the scale itself does not decide to which of these two factors the greater attention should be given. Even when the scale is placed in the hands of the teacher, these questions of relative value must still be decided. However, in this particular respect the scales themselves will work out their own salvation, for, by a consensus of expert opinion, it will be possible to decide for any particular grade the amount of speed that should be required as well as the degree of accuracy.

Another point against which school systems must carefully guard themselves, when these scales and standards are introduced, will be a tendency for schools to overlook those factors which do not admit of measurement by

such objective scales. This danger will gradually be eliminated as time goes on and as further scales for the measurement of school products are worked out. In the meantime, merely because only certain abilities at present admit of measurement, the school must not overlook subjects and factors which as yet do not admit of such quantitative estimation. In particular it must not fail to take into consideration such factors as the personal character of the teacher, the moral atmosphere of the school, and other spiritual values which, like life, beauty and happiness, are, to say the least, difficult fields for quantitative analysis. Such spiritual values in schools are of the greatest importance; to overlook or underestimate this fact would indicate a profound lack of sense of relative values. Even statisticians remember these things. But because we cannot estimate spiritual values, it is no reason why we should not measure values in those realms which admit of measurement. No science would have evolved, if it had not in its beginning confined itself to a limited field, and left large parts of the subject for the future. Furthermore, there is very strong *a priori* evidence to suggest that there is a close correlation existing between spiritual values and the values which these scales measure. If in the things we can measure it can be shown that the work is inadequate, there is every reason to believe that in the region of spiritual values there are shortcomings which escape our measuring rod. Certainly low objective values are no great argument for high spiritual values!

**The Future of Educational Measurement.** Many of these tests need criticism and revision, and such questions as their fairness and practicability can be answered only by the teachers who use them. For this reason the authors have refrained from any detailed consideration of the shortcomings of the individual scales. But the time spent upon their application will accomplish a twofold purpose:

It will improve the scales themselves; and it will give to every teacher who employs them a quantitative point of view which is sadly lacking in the schools, for many questions of school procedure do not admit of being answered by a mere affirmative or negative — the answer is found in the quantitative measurement. The Director of Reference and Research of the Department of Education of the City of New York says: "There could be no better exercise for a teachers' seminar than a series of discussions on some selected tests that would invite the independent judgment and criticism of intelligent teachers."

It is dangerous to forecast, especially when a subject is in its infancy, but there is every reason to believe that the application of the scientific method and the logic of statistics to educational problems will slowly revolutionize the method of education, even on its philosophical side. Moreover, in certain branches it will raise the study of education to the level of an exact science, thereby winning the respect of the scientific world for a subject whose low standards of proof and loose methods in the past have been responsible for the stigma which attaches to the study of education as an academic subject in the school and college curriculum.

### EXERCISES

1. When we are told that a child is "poor" in arithmetic, what is implied by this statement? How may we use the scales described to discover the point at which, and the extent to which, the individual is below standard?

2. How may the norms established for the scales actually confirm a school in lax teaching methods? How could this evil be prevented?

3. What other scales would be useful in the classroom?

4. How would you start to construct a rough objective scale for measuring (a) moral judgment, (b) æsthetic appreciation, and (c) humor?

5. How will the norms established by the use of these scales help greatly in settling the question of time distribution in the schedule?

6. Why is a single survey of a school of limited value? What are the advantages of measuring the quality of the work every half year?

7. How would you show a class the rate at which it was improving, from month to month, in order to accelerate its progress in (a) spelling, (b) writing, (c) reading?

8. How would you proceed to compare two different methods of teaching spelling by means of the objective scales? Enumerate the dangers and show how you would avoid them?

9. It is sometimes said, "These scales do not measure the most important work of the school, therefore they are of little avail." How would you meet this criticism?

10. How would you conduct, in a small city system, a general survey of the quality of the work done in the common subjects of the curriculum?

## APPENDIX

### SOURCES OF THE SCALES

The sources, from which a full account of each of the scales can be obtained, are given below.

COURTIS, S. A.

*A Manual of Instructions for Giving and Scoring the Courtis Standard Tests.* (75 cents.) 82 Eliot Street, Detroit.

This manual also includes the *Courtis Handwriting and Reading Scales*.

The standard blanks for any of the above tests, together with full directions for administration and scoring of the test, may be obtained from Mr. S. A. Courtis at the above address.

THORNDIKE, E. L.

*Handwriting.* Teachers College Record, 11: No. 2. 1910. (30 cents.) Publication Bureau, Teachers College, New York City.

Separate copies of the scale can also be secured (5 cents).

AYRES, L. P.

*A Scale for Measuring the Quality of Handwriting of School Children.* (5 cents.) Russell Sage Foundation, New York City.

THORNDIKE, E. L.

*The Measurement of Ability in Reading.* Teachers College Record, 15: No. 4, 1914. (30 cents.) Publication Bureau, Teachers College, New York City.

The standard blanks used in the Thorndike Tests may be procured in any quantity from the above address.

STARCH, D.

*The Measurement of Efficiency in Reading.* Journal of Educational Psychology, January, 1915. (30 cents.)  
Warwick and York, Inc., Baltimore.

The standard blanks for the administration of the test may be obtained, in any quantity, from the author, Dr. Daniel Starch, University of Wisconsin.

BUCKINGHAM, B. R.

*Spelling Ability: Its Measurement and Distribution.* (95 cents.) Publication Bureau, Teachers College, New York City.

STARCH, D.

*The Measurement of Efficiency in Spelling.* Journal of Educational Psychology, March, 1915. (30 cents.)  
Warwick and York, Inc., Baltimore.

AYRES, L. P.

*A Measuring Scale for Ability in Spelling.* (30 cents.)  
Russell Sage Foundation, New York City.

HILLEGAS, M. B.

*A Scale for the Measurement of Quality in English Composition by Young People.* Teachers College Record, 13: No. 4. 1912. (30 cents.) Publication Bureau, Teachers College, New York City.

BALLOU, F. W.

*Scales for the Measurement of English Composition.* (40 cents.) The University Press, Harvard University, Cambridge, Mass.

TRABUE, M. R.

*Completion Test Language Scales.* (\$1.15.) Publication Bureau, Teachers College, New York City.

The scales described, together with the Practice Sheet, may be purchased in any quantity from the above address.

THORNDIKE, E. L.

*The Measurement of Achievement in Drawing.* Teachers College Record, 14: No. 5. 1913. (30 cents.) Publication Bureau, Teachers College, New York City.

WOODY, C.

*Measurements of Some Achievements in Arithmetic.* (95 cents.) Publication Bureau, Teachers College, New York City.

The standard blanks for the administration of these tests may be procured in any quantity from the above address.

## BOOKS FOR FURTHER REFERENCE

### General

STARCH, D.

*Educational Measurements.* The Macmillan Company. (\$1.25.)

*Teachers' Year Book of Educational References.* Publications No. 6 and No. 14. Department of Education, City of New York.

Both the above books give very adequate bibliographies.

### Application of Scientific Measurement to a School Survey

JUDD, C. H.

*Measuring the Work of the Public Schools.* (50 cents.) Survey Committee of the Cleveland Foundation, Cleveland, Ohio.







UNIVERSAL  
LIBRARY



130 985

UNIVERSAL  
LIBRARY